2015

Taming the hashtag: universal sentiment, SPEQ-ing the truth, and structured opinion in social media

Ian La Vie
Iowa State University

Taming the hashtag: Universal sentiment, SPEQ-ing the truth, and structured opinion in social media

by

Erin Mikel Phillips

A dissertation submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Human Computer Interaction (HCI)

Program of Study Committee:

Eric Abbott, Major Professor

Jacob Groshek

Jan Boyles

Mack Shelley

Alex Tuckness

Iowa State University

Ames, Iowa

2015

DEDICATION

To the great ones in my life who reached down to help me grow: Paul and Arline Phillips, who gave me life and love; William LaChapell, who gave me belief and the courage to dream; Orville Elseth and Bruce Blom, who gave me opportunity and vision; and, Joanne Maloney, who loved me as a son and instilled in me a passion for excellence.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

My deepest and most heartfelt thank you goes to Dr. Eric Abbott, my major professor. Dr. Abbott's enthusiasm and relevant expertise have made a challenging and personally costly Ph.D. journey extremely rewarding.

A special thank you goes to my committee members: Dr. Alex Tuckness [Political Science], Dr. Jacob Groshek [Voting and Media], Dr. Jan Boyles [Journalism], Dr. Vassant Havovar [Computer Science and Semantics], Dr. Mack Shelley [Political Science and Statistics], and Dr. Richard Freed (emeritus) [Rhetoric]. With such a multidisciplinary topic, each member played an essential role. Some lent expertise from their respective fields relevant to my research. Some encouraged me personally. All provided excellent and sustained support amidst both the grand and the mundane activities inherent in doctoral pursuits.

I am also thankful for the patient affection of my wife and children. Robin, Natasia, Joshua, Sasha, Noah, Nathanael, Elianna, Paul, Katerina, and Thomas—thank you.

ABSTRACT

Opinions are valuable, and with the advent of social media, plentiful. Opinions are not always intelligible, however. Therefore, many of the views of social media users are ignored. This dissertation seeks to confront the challenges associated with opinion mining and sentiment analysis by investigating three aspects of opinion expression and consumption in social media. The universality of opinion itself is explored through an innovative application of social science research in survey construction, semantic distance analysis, and corpus linguistics. Results include a universal taxonomy of 18 sentiment types shown to be portable across 15 languages. The universality of opinion processing is explored through a qualitative meta-synthesis (QMS) analysis of social psychology, opinion mining and sentiment analysis, and voting systems scholarship. Results include a comprehensive theoretical model of opinion processing: the States, Processes, Effects, and Quality (SPEQ) model for opinion mining and sentiment analysis. SPEQ defines seven states of opinion, six processes which govern the transitions between those states and five quality and integrity measures for the evaluation of those processes. Lastly, the concept of a structured opinion syntax is explored. Despite strong resentment to symbolic representations of meaning by subjects, learning and priming effects for both the encoding and decoding of structured opinion support the contention that such a syntax could be developed and used. Many future directions for research are presented for each aspect of opinion investigated.

CHAPTER 1. GENERAL INTRODUCTION

"Opinion is a powerful, bould, and unmeasureable party."

— De Montaigne (1580), "That the taste of goods or evils doth greatly depend on the opinion we have of them."

"The public is almost always ahead of its governmental leaders. This statement has been made many times, and it can be supported by an overwhelming volume of evidence amassed during these two decades on nearly every conceivable issue-political, social and economic."

— Gallup (1957), "Wisdom of the public."

"We suggest general performance requirements for end-to-end verifiable elections and not on [verifiable] voting systems: we care if the election outcome accurately reflects the intentions of the voters, regardless of whether the voting equipment is "correct" or not . . . it is ultimately the election that is checked, not just the equipment."

— Popoveniuc et al. (2010), "Performance requirements for end-to-end verifiable elections."

1.1 Introduction

Opinions matter. Governments, organizations, societies, families, and individuals that value the opinions of the relevant populations—are stronger for it. Those that hide from, or pro-actively stifle opinions eventually find they are unable to sustain the effort. The quotes from De Montaigne (1580), Gallup (1957), and Popoveniuc, Kelsey, Regenscheid & Vora (2010) show us that there is something visceral, ennobling, and sustaining about opinion. Forgetting the formality of this document for a moment, this investigator confesses to having developed a certain awe, even affection, for the concept of opinion through the doctoral process. The ability to form, faithfully express, and reliably consume opinion is a sustaining capability of our species. With it, there is hope—without it, there is hopelessness.

Social media has created a new era for opinion expression. The proliferation of social media platforms and devices which enable their use has enabled billions of individuals to generate even more opinions. Opinion expression is reaching levels which are hard to conceptualize. A profound asymmetry has resulted, however. There are not enough humans to consume all of these opinions, and automated opinion decoding is (still) in its infancy. Opinion mining and sentiment analysis research is mired in dealing with the vagaries and complexities of language through ever more elaborate and sophisticated algorithms and lexicons. In his extensive review of opinion mining and sentiment analysis scholarship over the last decade, Liu (2012, p. 13) inventoried numerous challenges in opinion mining and sentiment analysis, then summarized the state of affairs this way: "These issues all present major challenges . . . in fact, these are just some of the difficult problems."

The scope and approach of each of the three papers which make up this dissertation are focused on understanding and potentially reducing this asymmetry by confronting the major challenges facing opinion mining and sentiment analysis today. It is hoped that in the long run that this or similar research will shift the rhetoric around opinion expression in social media from "everyone is talking" toward "everyone is being heard."

This dissertation pursues a comprehensive review of recent literature on opinion mining and the related disciplines of social science, corpus linguistics, social psychology, and voting systems. Built on this foundation of scholarly reflection are three papers which present innovative experiments and "daring generalizations" (Albig, 1957), including a new theoretical model of opinion.

1.2 Dissertation Organization

Three papers are presented in this dissertation, preceded by the this introduction (Chapter 1), and summarized subsequently with conclusions and recommendations for future research in Chapter 5. Three distinct but related lines of inquiry motivate the three papers presented.

Is opinion expression universal? If so, how? The first paper, "Cross-cultural opinion parity: is there a universal taxonomy of sentiment types?", addresses these questions and is presented in Chapter 2. This paper summarizes research which integrates research in social science survey construction, semantic distance analysis, and corpus linguistics, in an attempt to develop then validate a taxonomy of sentiment types which is portable across languages.

Is opinion processing universal? If so, how? The second paper, "SPEQ-ing the truth: the states, processes, effects, and quality model for opinion mining and sentiment analysis" addresses these questions and is presented in Chapter 3. This paper summarizes research which integrates research in opinion mining and sentiment analysis, social psychology, and voting systems, to define a comprehensive lifecycle of opinion.

Is there a way to leverage both the universality of opinion expression and the universality of opinion processing to create a more reliable form of encoding and decoding opinions? If so, how? The last paper, "foo#that and #this+++: a structured sentiment usage study", addresses these questions and is presented in Chapter 4. This paper provides a cross-cutting look at how individuals respond to an opinion encoding syntax. The definition of a structured opinion encoding syntax relies on the universal nature of opinion explored in Chapter 2 and enables a more reliable conveyance of opinion through the universal processes developed in Chapter 3.

CHAPTER 2. CROSS-CULTURAL OPINION PARITY: IS THERE A UNIVERSAL TAXONOMY OF SENTIMENT TYPES?

Accepted for inclusion in proceedings of KONVENS 2014.

Erin Mikel Phillips[1]

Abstract. Opinion mining and sentiment analysis research has focused on 2-state (-1, +1) or 3-state (-1, 0, +1) representations of sentiment; however, the semantics of opinion are more complex. Moreover, research on opinion mining in social media has tended to be monolingual while social media usage is known to be a global phenomenon. These challenges are due in part to the absence of a sentiment classification scheme which is portable across languages. In this paper, a cross-language taxonomy of 18 sentiment types is developed through an innovative application of social science research in survey construction, semantic distance analysis, and corpus linguistics. This taxonomy, the "Universal 18" (U18), was tested for rank-order consistency across 15 languages in two social media corpora containing 400M documents. Results show U18 usage to be consistent across the languages studied. Moreover, because the two corpora used are aged three years apart, there is some evidence that these findings are reliable retrospectively and durable prospectively. These findings suggest that social media authors express types of sentiment in similar proportions, regardless of the language used and provide a basis for expanding opinion mining beyond polarity detection. Future directions for research include the development of a universal sentiment syntax and the use of sentiment trace density as a SPAM detection criterion.

[1]Primary researcher and author.

2.1 Introduction

The concept of opinion, an awareness of its value, and an appreciation of its elusiveness is as old as the human race itself. As De Montaigne (1580, p. 254) wrote, “Opinion is a powerful, bould, and unmeasureable party.” De Montaigne’s maxim of the perpetually latent public sentiment persists to this day. The numerous and increasing computational and linguistic approaches to sentiment analysis inventoried by Pang & Lee (2008) and Liu (2012) are evidence of that. The advent of democracy, continuous polling, and opinion mining of social media content have indeed expanded the channels through which public opinion is accessible; however, significant barriers remain. Modern elections are arguably the most structured and controlled process for capturing public sentiment yet implemented. These, too, suffer from fundamental challenges of voter access and vote encoding and decoding accuracy (MIT, 2001; Stenbro, 2010). Access and accuracy are two of the prominent challenges facing voting systems scholars today. Opinion mining researchers—whose voter rolls and votes are of a vastly inferior quality to those found in elections, also face substantial challenges regarding access and accuracy.

2.1.1 Problem

We are well into the second decade of what Pang & Lee (2008, p. 7) called “the sentiment analysis and opinion mining ... land rush.” However, opinion mining research is still predominantly both using single-language corpora—the access problem, and using oversimplified opinion semantics—the accuracy problem. Definitions of key terms are discussed in more detail in Section 2.2.2.

The Access Problem: Limited Linguistic Diversity

The term access is used within to bridge the disciplines of voting systems scholarship and opinion mining research. In the context of voting systems scholarship, access refers to the opinion encoding potential of a relevant population: 100% access would equate to 100% of those that choose to vote, can vote. Opinion mining research largely ignores the question of opinion encoding potential, and focuses on the more immediate and tractable problem of trying to capture the opinion decoding potential of the opinions which have been encoded in text. However, the opinion decoding potential of social media includes expressions in many languages while the research itself is principally monolingual.

The lack of linguistic diversity in both natural language processing research and opinion mining scholarship is a problem which has been known for some time.Abbasi et al. (2008, p. 9) states, "most studies have focused on sentiment classification of a single language." Surveys of opinion mining research as recent as 2012 have described cross-language approaches as recent extensions to the field. The following quote from Liu (2012) is included because the reference is recent, authoritative, and highly relevant to the focus of this research.

> "Recently, several extensions to this research have also appeared, most notably, cross-domain sentiment classification (or domain adaptation) and cross-language sentiment classification" (p. 31)

It is important to note that the problem is not limited to opinion mining research. Indurkhya & Damerau (2012, p. 25) provides this assessment of rule-based approaches in natural language research: "[rule-based] systems are usually developed for a single language... as a result they are not portable to other natural languages."

Causes. Communication between parties, even when face to face, presents challenges for both sender and receiver (Shannon, 1948). If the communication between parties is written, the challenges increase dramatically. As Indurkhya & Damerau (2012, p. 9) points out, "writing systems often amplify ambiguities [in language]." Idiom, satire, sarcasm, bad grammar, and human error can each create significant or even unsolvable problems for the researcher trying to develop a mechanism which can render a reliable interpretation of an opinion author's intent. The most obvious cause, then, of lack of linguistic diversity is that extracting the author's meaning from written free text in any single language is a significant challenge.

Though an indirect cause, it also seems plausible that the rush to identify practical applications through empirical methods has caused a lack of qualitative research of the kind which yielded powerful theoretical generalizations such as Shannon (1948), Rogers (1976), or Winston (1998). Moreover, the foundations of opinion mining research are primarily empirical studies, such as Novak et al. (2000), Pang et al. (2002), Tang & Liu (2005), and others. These researchers set the stage for an explosion of empirical approaches, with the mountains of available social media content providing the fuel. Even the most basic type of sentiment detection, polarity detection, involves the intricate application of various quantitative approaches, each of which can explore interesting nuances in the use of a single language—typically within a single corpus.

Consequences. As a result of what could be described as a type of narrow empiricism, our understanding of how diverse populations express opinion through social media may be limited. When a natural language processing or opinion mining research study uses text in a single language as the experimental data set, the ability to generalize from the study findings to other languages is weakened substantially. In light of the present and expanding global proliferation of social media, the lack of linguistic diversity has kept opinion mining research in social media from keeping pace with opinion expression in social media. A widely accepted framework for cross-language analysis of sentiment has yet to emerge. An example of a recent study may help explain the situation.

The recent study by Zheng et al. (2014) examined 10,000 reviews across 4 topic categories of reviews (hotel, restaurant, mp3, camera) in English were analyzed using an elaborate heuristic based upon parts of speech relationships. 2000 reviews were selected as a "gold standard" and evaluated by three graduate students to determine topic(s) and aspect(s) intended by the review authors. The resulting sentiment topic and aspect classification accuracy were 60-70%. No attempt was made to evaluate the actual opinion itself. The scope only included the identification of the relevant aspect or topic. A number of variations on the core algorithm are investigated in the paper; however, no attempt was made to generalize to other languages.

The research presented in Zheng et al. (2014) is both creative and rigorously approached. However, the methodological approach yields little if anything which can be used to foster the type of cross-cultural or cross-language scholarship needed for useful generalization to the global phenomena of social media. The problem of lack of linguistic diversity can be summarized as follows:

Problem 2.1. There exists a gap between the monolingual character of opinion mining research in social media and the multi-lingual phenomena of opinion expression in social media.

The Accuracy Problem: Oversimplified Opinion Semantics

The term "accuracy" has varied and nuanced meanings in the context of both voting systems scholarship and opinion mining research—but its essence in both realms is truth. In voting systems scholarship (Hosp & Vora, 2008), accuracy (also called integrity) refers to the congruence between a voter's intent and the encoded vote. This usage is highly relevant to opinion mining scholarship. As Problem 2.1 discussed the opinion encoding potential of a population, in this section the opinion encoding potential of an individual vote is discussed. In the context of voting systems scholarship, "accuracy" refers to the opinion encoding potential of a voter's intent. 100% accuracy would equate to 100% of meaning which the voter intended to convey through voting, and which can be conveyed through voting, was encoded in the vote.

With opinion mining research, however, the "voting machine" is often 160 characters of free text in a Twitter post or 10,000 words of free text in a blog post. Because of the challenges discussed above regarding extracting meaning, opinion mining research largely ignores the question of opinion encoding potential in a particular opinion. Instead, researchers focus on the most tractable problem of trying to classify an opinion encoded in text using an abstraction, such as positive or negative. The core elements of opinion semantics, as defined by FrameNET (Baker et al., 1998), are typically either assumed to be held constant or ignored. These include a cognizer, a way-of-thinking or private state, topic, domain [aspect], constancy, evidence, manner, role, and time. Definitions of key concepts are discussed in more detail in Section 2.2.2.

While the use of these abstractions is a natural step in the process of developing knowledge, it is not clear from the literature of the field that there is a general recognition of the semantic gulf between "opinion mining accuracy" using social media text and "accurately mining opinions" from social media text.

Thousands of articles have been written presenting various algorithms and approaches to identifying bias or polarity in text. While focusing on the direction of opinion in opinion-laden social media content, many of these studies are executed within the operational framework discussed above from Zheng et al. (2014). Studies following a similar to approach to Zheng et al. (2014) include Hu & Liu (2004), Kim & Hovy (2004), Xu, Wong & Xia (2007b), Go, Huang & Bhayani (2009), Pak & Paroubek (2010), and Zhai, Liu, Wang, Xu & Jia (2012). The use of these abstractions is a natural step in the process of developing knowledge. However, it is not clear from the literature of the field that there is a general recognition of the gap between "opinion mining accuracy" and "accurately mining opinions."

An example can illustrate the oversimplification of opinion semantics. Pak & Paroubek (2010, p. 1) proposes to, "show how to use Twitter as a corpus for sentiment analysis and opinion mining. . . [to explore questions such as] 'What do people think about our product (service, company etc.)?'" As in Zheng et al. (2014) and many other opinion mining studies, Pak & Paroubek (2010) manually annotates a small corpus of texts. The following encoding scheme is used: positive ("texts containing positive emotions"), negative ("texts containing negative emotions"), and neutral ("objective texts that only state a fact"). Consider an example micro-blog post (Example 2.1) taken from Pak & Paroubek (2010). Is Example 2.1 positive, negative, or neutral? Does it contain an opinion or perhaps more than one? What does it say about Obama? Or, Chicago? Or, "the games?"

Example 2.1. funkeybrewster: @redeyechicago I think Obama's visit might've sealed the victory for Chicago. Hopefully the games mean good things for the city.

The methodology of human annotated corpora is not the problem, and neither is the use of some mathematical abstraction from the author's intent. The problem is the lack of clarity around what is being measured. The encoding scheme used by Pak & Paroubek (2010) is not wrong, but it's not opinion either. **`@redeyechicago`** is the cognizer or author of the statement, but what is the 'way of thinking'? What is the topic? What is the domain (or aspect) of comparison? What is the manner (or strength) of the opinion? These questions are unaddressed by the study because the study does not reference the semantic definition of opinion. Specifically, the words "semantic" or "meaning" or even "hypothesis" do not appear at all in the paper. The term sentiment analysis is sometimes used to lower the semantic expectations around the polarity detection task. However, as Liu (2012, p. 7) points out, "*Sentiment analysis*, also called *opinion mining* ... They basically represent the same field of study."

Causes. A thorough discussion of the root causes for the disconnect between full opinion semantics and opinion mining research is beyond the scope of this inquiry. Moreover, the complexities discussed above in Problem 2.1 which motivate researchers to pursue opinion mining research using single-language corpora, are also relevant to the problem of accuracy. That the two or three-state polarity model of opinion falls short of the known semantics of opinion is obvious when mentioned. However, the polarity detection task is so tangible and easily interpreted, that introducing a more complex representation may seem to be of little value. A more subtle and linguistic cause may lie in the $+1$ and -1 representation of opinion which has become a common visual metaphor for the expression and reporting of opinion. The meaning of these symbols is closely aligned with the concept of sentiment polarity in the literature.

Consequences. The use of oversimplified opinion semantics in opinion mining research has created large contradictions in the literature. On the one-hand studies report high levels of accuracy and broad claims of practical application. Pak & Paroubek (2010, p. 1325) summarize their results: "We can obtain a very high accuracy ... if we use our classifier for the sentiment search engine, the outputted results will be very accurate." On the other hand, published surveys of opinion mining literature indicate that the practical value polarity of detection studies is limited because the useful information available from polarity detection algorithms is limited. As shown above, what do $+1$ and -1 mean? As Liu (2012, p. 14) states, "classifying the sentiment or subjectivity expressed in documents or sentences [using polarity] is insufficient for most real-life applications . . . practical applications often demand more in-depth and fine-grained analysis."

The overemphasis on empirical abstractions has left opinion mining research in something of a stagnant posture in terms of practical applications. The field of opinion mining has not advanced as rapidly as it might if more effort was given to exploring the full semantics of opinion—which might in turn lead to richer and more fruitful theoretical generalizations. The second major problem this research seeks to address can be stated as follows:

Problem 2.2. The operational models used in many opinion mining research studies explore abstractions of opinion, rather than the semantics of opinion.

While an aside, it is worth mentioning that the situation facing opinion mining research appears similar to that facing Public Opinion researchers early in the development of that field. See Albig (1957) for a lengthy criticism of the state of Public Opinion Research after two decades of polling and a "land rush" of sorts for the empirical analysis of polling data.

2.1.2 Purpose

The general purpose of this study is to explore options for expanding the avenues of approach and discourse around opinion mining research, especially in the direction of improving access and accuracy as described in Section 2.1.1. In doing so, it is hoped that the outcomes will support the proposition that Problem 2.1 (access) and Problem 2.2 (accuracy) are, in fact, real problems. Additionally, some innovative ways to attack the problems are demonstrated. The following research questions set the boundaries for this inquiry. Because of an operational dependency, the research question relating to accuracy is mentioned first.

Research Question 2.1. *What are the most common types of opinion expressed in English social media?*

The specific purpose of Research Question 2.1 is to extend opinion mining to include a taxonomy of sentiment types as a new dimension of analysis.

Research Question 2.2. *How portable is the list of opinion types derived from Research Question 2.1 for English, to opinions in social media expressed using languages other than English?*

2.1.3 Approach

This section provides a summary of the approach used to accomplish the purpose of this paper described above in Section 2.1.2. As indicated by Research Questions 2.1 and 2.2, there are two research tasks that work together to enable the research described in this paper. Within the discussion on methodology in Section 2.3, these two tasks are described in formal logic in Section 2.3.1 and specific procedural steps in Section 2.3.3. The approach used for exploring Research Question 2.1 is innovative. It starts with the social science scholarship on attitude scales of Likert (1932) and Vagias (2006). Then, it applies the lexical-semantic research of Miller et al. (1990), Wiebe & Mihalcea (2006), and Piasecki et al. (2010).

In summary, an inventory of 37 of the most commonly used social science attitude scales, or, Likert-scales (Likert, 1932), found in survey design is used. This list of scales was developed by Vagias (2006). These 37 scales were put through a lexical normalization process to reduce the original set to a set of semantically orthogonal scales. The normalization process used in this research relies on the revolutionary work in semantics by Miller et al. (1990). The end-result of this first task is a set of 18 semantically distinct attitude or sentiment scales (called, U18, herein) based upon social science scholarship on survey construction. This taxonomy of the most commonly used sentiment scales was vetted to have minimal semantic overlap. It serves as a primary input to Research Question 2.2, namely, whether such a taxonomy can be useful for classifying the sentiment of social media users beyond those using English.

Recall that Research Question 2.2 involves evaluating the cross-language character of the taxonomy of sentiment types developed as an outcome of Research Question 2.1. The approach used here for exploring Research Question 2.2 is similar to the work done by others (Tokuhisa et al., 2008; Wan, 2008; Kamińska & Pelikant, 2012). However, some important innovations are worth noting here. Kamińska & Pelikant (2012) used manual annotation to encode voice wave features as markers to Plutchik's model of 8 emotions. Tokuhisa et al. (2008) developed a list of words and phrases which are lexical markers of 10 possible emotional states of the writer.

The use of human annotators is a common and useful method of developing taxonomies which embody heuristics too complex to parameterize. In the case of this research, it was determined that sufficient scholarship on subjectivity supported a more automated and objective approach. It is well established that the presence of a scaling adverb or adjective is strong semantic marker for sentiment. Such a marker is a stronger indication of sentiment than the verb or noun alone (Breck et al., 2007; Tang et al., 2008). It follows then that an inventory of adverbial modifiers would make an excellent list of opinion markers. A searchable corpus was used to extract a list of the most commonly used adverbial and adjectival scaling modifiers for each of the resulting 18 sentiment types. The Corpus of American Contemporary English (COCA) (Davies, 2009) was the searchable corpus used. The result of this process was 960 scaling phrases (called, "sentiment traces," hereafter) in English, approximately 50 for each of the 18 sentiment types.

The scope of Research Question 2.2 includes a cross-language analysis, so the inventory of 960 sentiment types was translated into 14 additional languages using automated translation. The manner of translation is similar to that used by Wan (2008).

Lastly, opinion mining and sentiment analysis research frequently relies on large social corpora. These provide sufficient textual discourse to obtain useful results. Mishne (2005) used an extensive blog corpus for an opinion mining and sentiment analysis task which included the identification of the mood of the author. NLP classifiers were used in Ptaszynski et al. (2012) to annotate a 5B word blog corpus of Japanese blogs for both subjectivity and emotion expressed.

This research uses two blog corpora (Burton et al., 2009, 2011) from the same source, separated in time by a period of approximately three years. These corpora combined to hold approximately 400M social media documents in 30+ languages. Only English and the top 14 other non-English languages were used in this research because those languages constitute 97.4% of the documents. The final step in the approach to this study was to examine the frequency of occurrence of each of the 960 sentiment traces. This examination was done for all 960 sentiment traces in each of the 15 languages. The rank order correlation across the languages was used to evaluate the portability of U18.

If the rank order correlation is high between languages, then that would indicate that social media users of different languages tend to express opinions using the sentiment types with similar relative frequencies. For example, if Russian social media users and Dutch social media users express subjective statements about "quality" in a similar relative frequency, then "quality" can be said to be a portable, or cross-language sentiment type.

If the rank order correlation is low between languages, then that would indicate that social media users of those languages are not expressing opinions similarly. For example, if Spanish social media users express subjective statements about “agreement” in much lower relative frequencies than Arabic social media users, then “agreement” is not a portable or cross-language sentiment type.

2.1.4 Significance

This work is significant in terms of its design, and in terms of its outcomes and opportunities for future research.

Design. The experimental design is significant because it is the first study to combine social science survey design scholarship with opinion mining scholarship. This approach grounds the U18 taxonomy of sentiment types in social science, as opposed to manually developed taxonomies derived from a plurality of human annotators.

The experimental design is also significant because it is repeatable. Few, if any, of the opinion mining and sentiment analysis studies found in this course of this research, are free of human annotation. The automation in this paper includes the establishment of a semantically orthogonal taxonomy of 18 sentiment types. The development of the inventory of 960 sentiment traces was done through automated searches. The translation of the sentiment traces into 14 non-English languages was done using automated translation. The preparation of the corpora according to defined rules using automated programs which applied those rules. Finally, the construction of the regular expressions and the subsequent scanning of the corpora for sentiment type frequency values were both done through automated processes.

Future research. This research is also significant because of the new avenues of approach to opinion mining and sentiment analysis it opens. As discussed throughout the balance of this paper, the tendency in related research is to focus on polarity detection, $+1$ and -1, as a measure of opinion. This research leverages social science scholarship to define new dimensions of analysis. Moreover, by introducing the concepts of a "sentiment type" and "sentiment trace," this research highlights weaknesses in the canonical definition of opinion which lacks these elements.

Lastly, the affirmative outcomes of this research show that social media users which write using English, or the other 14 languages studied, tend to share opinions along similar dimensions of comparison. This finding is revolutionary, in the sense that it offers a first empirical glimpse into universality of opinion expression within the global social media experience. Additional details are provided in Section 2.5.

2.2 Background

This section contains a review of relevant scholarship. The framework for this literature review is the semantic definition of opinion provided by (Baker et al., 1998). There are some challenges with the literature review for this research. Within the field of opinion mining and sentiment analysis, published research is dominated by empirical studies which evaluate algorithmic approaches to extracting meaning from free text. Therefore, the number of interdisciplinary and theoretical works available for review is small, and other related disciplines are therefore consulted for important concepts. These other disciplines include psychology, social psychology, and linguistics.

2.2.1 Diligence

The execution of the literature review is modeled after the concept explication method described in Chaffee (1991). However, because the scope of this research spans multiple disparate disciplines, some effort was made to develop quantitative methods adapted from corpus linguistics to identify sources of influence in relevant papers.

Qualitative Methods

The qualitative method used for this literature review did not involve specific literature review indices. A previously compiled definitive list of established sources which would undergird this research directly could not be found.

However, scholar.google.com was also used extensively to identify sources of scholarship and to trace reliance relationships. Though no indices were located, two comprehensive surveys of opinion mining scholarship serve as the milestone markers for much of the opinion mining scholarship today:Pang & Lee (2008) and Liu (2012).

The Pang & Lee (2008) volume is an excellent summary of the formative scholarship in the field. With 332 references, most major topics and approaches discussed in opinion mining research from 2000 to 2007 are covered. Pang & Lee (2008, p. 1) state in the introduction, "This survey covers techniques and approaches that promise to directly enable opinion-oriented systems."

The Liu (2012) volume is the most recent comprehensive summary of opinion mining research yet published. With 403 references and a substantial amount of analysis and synthesis, the Liu (2012) volume reads a little more like an opinion mining research manifesto than an anthology.

> "the goal of this book is to give an in-depth introduction to [opinion mining and sentiment analysis] and to present a comprehensive sur-

> vey of all important research topics and the latest developments in the field ... bridging the unstructured and structured worlds and facilitating qualitative and quantitative analysis of opinions. This is crucial for practical applications." (p. 5)

Because of the comprehensive nature of these volumes, they were consulted extensively in the identification of other supporting resources.

Quantitative Methods

The diversity of subject matter relevant to this research, combined with the inexperience of the primary investigator, prompted the development of specialized content analysis tools to aid in the literature review. These tools were written using a parser generator developed by this researcher from an activity unrelated to this research. The purpose of these tools was twofold: enable bibliographical analysis of the literature; and, enable a query facility capable of scanning for citations whose surrounding lexical context seems relevant to a particular n-gram. Details of the CiteScan tool's design are provided in Appendix A.

2.2.2 Concepts

The following topics are foundational to this research. Each is presented using existing scholarship for definition and usage in research.

Semantic Frame. A semantic frame is a unit of meaning, a structured representation of human knowledge about the relationship between lexical or syntactic elements and other semantic frames. The FrameNET (Baker et al., 1998) project maintains a database of semantic frames for thousands of semantic domains, including emotion and cognition.

> "includes hand-tagged semantic annotations of example sentences extracted from large text corpora ... [using] semantic patterns they ex-

> emplify by lexicographers and linguists. The primary emphasis of the project therefore is the encoding, by humans, of semantic knowledge in machine-readable form." (p. 86)

The components of a semantic frame are either core (or required) elements or non-core (or optional) elements. The concept of a semantic frame is important to this research because it sets the boundaries and defines the particulars of what opinion means. The term opinion is defined in some detail within FrameNET. The semantic frames within FrameNET are maintained by lexicographers, linguists, and experts in the domains of emotion and cognition.

Private States. The concept of private states is important in social psychology and social science scholarship because they represent a person's beliefs and desires which form the attitudes which influence human action Reisenzein (2009). Quirk et al. (1985) is frequently cited on this point. Quirk establishes a criterion for private states as, "a state that is not open to objective observation or verification." Wiebe & Deng (2014, p. 5) defines private state as, "[an] attitude held by a source toward (optionally) a target," which contains the core elements of the semantic frame for opinion.

Subjectivity. Private states remain private unless expressed in some way. Linguistic expression is one way in which a projection of a person's private state is made available to others. As Wiebe & Deng (2014, p. 5) declares, "Subjectivity is the linguistic expression of private states." The concept of subjectivity is important in this inquiry because subjective statements presuppose reliance on a private state—which is the essential characteristic of an opinion. The presence of a scaling adverb or adjective has been shown to be a reliable lexical cue to the presence of a subjective statement (Breck, Choi & Cardie, 2007; Tang & Liu, 2005).

Opinion. The concept of opinion is rooted in the concept of subjectivity. Post (1990) provides a legal definition of opinion which aligns itself with private states as being the essence of subjectivity and the defining characteristic of opinion.

> "[opinions] are not objectively verifiable or subject to empirical proof... for constitutional purposes the truth of certain kinds of statements — *opinions* — can only be determined by the free play of speech and counter-speech characteristic of the marketplace of ideas." (p. 656)

The semantic frame for opinion also relies on the existence of a private state of a person—or more precisely, a cognizer. Baker et al. (1998) presents the semantic frame for opinion as, "A *Cognizer*[core] holds a particular *Opinion*[core], which may be portrayed as being about a particular *Topic*[non-core]." There are also sub-elements within the semantic frame for opinion. The full semantic frame for opinion is comprised of the following elements: cognizer[core], opinion[core] (also, private state), topic, domain (or, aspect), constancy, evidence, manner, role, and time. Baker et al. (1998) defines the opinion element as, "The Cognizer's way of thinking, which is not necessarily generally accepted, and which is generally dependent on the Cognizer's point of view."

Lastly, within opinion mining research, a formalism for opinion has developed which resembles the FrameNET opinion frame. Liu (2010, p. 633) defines opinion in the context of opinion mining research as, "the quintuple of (object, feature, orientation, holder, time)."

Aspect. Baker et al. (1998) defines domain as, "The aspects of the Opinion (and its Topic, if any) which are under consideration." The FrameNET definition of domain to the opinion mining term aspect are linked in a clear and direct association. The concept of aspect is important to this research in the sense that an aspect is not a private state, but an ascribed feature of the entity which is the target of the opinion. The distinction is important for opinion mining research. Liu (2012, p. 58) tells us that, "a positive opinion document about the entity does not mean that the author has positive opinions about all aspects of the entity." Aspects, then, are not attributes of private states; they are attributes of the target entity.

Sentiment. In the field of social psychology, the concept of sentiment is closely linked to emotion or feelings. Reisenzein (2009, p. 221) declares that sentiments are, "conscious nonconceptual metarepresentations: they are feelings that represent to experiencers, in a nonconceptual way, important states and (impending) state changes in their core representation system."

Given the direct association of sentiment to feelings, it is worthwhile to check the semantic frame for feelings. Baker et al. (1998) defines the feelings frame as, "an *Experiencer* experiences an *Emotion* or is in an *Emotional state* . . . [and optionally] an *Evaluation* of the internal experiential state." In sentiment analysis and opinion mining literature, sentiment is typically defined as Wilson et al. (2005, p. 347) defines it, "[sentiments are] positive and negative opinions, emotions, and evaluations."

Sentiment Scale. The formalism for classifying sentiment is rooted in social psychology, especially the work of Likert (1932). Likert (1932, p. 9) connects attitude or sentiment scales with opinion: "declarations of opinion and attitude are regarded as an indirect method of measuring dispositions." The result of this proposition was an approach to measuring sentiment that is often referred to as a Likert-scale (pronounced, 'lie-kurt').

As described in Edmondson (2005), Likert devised a formalism for survey question design which countered the "Thurstonian scaling technique" prevalent at that time. The Thurstonian method involved using a panel of experts to determine the favorableness of a series of statements relative to some questions or issue. The outcomes from these panels would be used to assign a baseline to the statements included in a survey, and the results of surveys would then be calculated using the panel's scaled values to quantify survey responses. The concept of a sentiment scale is critical to survey design scholarship (Conrad & Schober, 2007). The statistical character of any particular finding from a survey is predicated on the semantic orthogonality of the scale used.

With respect to sentiment scale, the number of them seems more limited than opinions themselves. Liu (2012, p. 88) observed that, "there seem to be an unlimited number of ways that people can use to express ... opinions." However, the number of types of sentiment (and their corresponding sentiment scale) does not appear to be unbounded in the same manner—that is, in the sense of everyday human experience. This statement makes some sense given the close relationship between emotion (an unanchored feeling) and sentiment (a feeling anchored to a private state—a meta-representation or projection of possible changes in a private state).

While the range of human emotional experience is extensive, the number of emotions seems limited. Ortony & Turner (1990, p. 315) exhaustively compared research on basic emotions and concluded, "not all of the variation in lists of basic emotions is real because the same emotion is often labeled differently by different researchers ... some theorists use the term anger and others the word rage while presumably referring to the same emotion."

Likewise, while the linguistic role of adverbial and adjectival anchors makes the list of sentiment-types theoretically infinite, the types of human experiences relative to changes in private states does not likely follow the same level of expansion. An inventory of social science Likert-type scale response anchors by Vagias

(2006) cataloged 37 sentiment-type scales—several of which were variations, such as 5 and 7-point scales for satisfaction.

Semantic Proximity. Semantic proximity is a powerful idea and the target of much research in Linguistics. Closely related to the concept of meaning, semantic proximity is an attempt to graph out relationships among lexical-semantic elements, so as to be able to determine how elements are related if at all.

Miller et al. (1990)introduced WordNet, a lexical-semantic database which enabled the connection between words and their dependents and derivatives. By providing the hypernym relation, WordNet establishes a hierarchy of semantics of words. Many have used the hypernym relations in WordNet in research related to this inquiry. Wiebe & Mihalcea (2006) showed that the hypernym relation can be used for disambiguation and classification. In an effort to improve the ability of linguists to visualize semantic relations, Piasecki et al. (2010) used the hypernym relation to draw connected graphs which let linguists explore the semantic relations between lexical units. The WordNet power of disambiguation is a foundational linguistic capability used in this research. Any effort to construct a taxonomy of sentiment types necessarily involves disambiguation of the terms.

2.2.3 Criticism

A fair amount of criticism has been discussed previously in Section 2.1.1. Some points remain, however, and will be covered in this section.

First, some work was done by Tokuhisa et al. (2008) to decompose sentiment polarity into particular emotions. However, this work focused on mapping lexical indicators of emotional state (i.e., fearful, sadness, anxiety) back to the polarity model of sentiment. No effort was made to develop a more granular classification for the "type" of sentiment being expressed.

By way of demonstration, do the following: express your opinion by answering the survey question in Example 2.2 derived from recent opinion mining scholarship:

Example 2.2. $+1$ or -1 ?

The nonsensical construction of Example 2.2 is not a mistake, but rather a look at opinion mining research measures from the viewpoint of social science and social psychology researchers.

While some liberties are being taken with the above device, the central criticism of this paper on opinion mining and sentiment analysis scholarship should be clear. The operational definitions of opinion and sentiment do not take advantage of the definitions developed over time by scholars in other disciplines. As Liu (2010, p. 44) states, "Knowing only [a sentence] positive or negative opinion, but not what entities/aspects the opinion is about, is of limited use."

2.2.4 Inferences

Sentiment Trace. The term "sentiment trace", introduced in this paper, refers to an adverbial or adjectival sentiment type exemplar which is later used to detect the presence of the corresponding sentiment type. For example, "disagree" and "also agree" are sentiment traces of the sentiment type "agreement" (see Table 2.3). The presence of a scaling adverb or adjective is a stronger indication of sentiment than the verb or noun alone (Breck et al., 2007; Tang et al., 2008). The terms "sentiment trace", and "sentiment trace scale", however, are not present in the literature—but can be inferred from both opinion semantics and social science research.

As described in Edmondson (2005), Likert devised a formalism for survey question design which countered the "Thurstonian scaling technique" prevalent at that time. The Thurstonian method involved using a panel of experts to determine the favorableness of a series of statements relative to some questions or issue. The outcomes from these panels would be used to assign a baseline to the statements included in a survey, and the results of surveys would then be calculated using the panel's scaled values to quantify survey responses. The problem addressed by Likert's early work with attitude classification methods mirrors the situation facing opinion mining research today and serves as a foundational concept and motivation for this study.

The theoretical framework consists of a series of concept definitions, a set of relationships between the concepts defined in previous scholarship and a set of new concepts and relationships proposed by this research.

Private states influence human actions, including writing. Written communication about externally verifiable phenomena is referred to in the literature as objective statements. Subjectivity in written communication, regardless of the form (i.e., lexical, grammatical, satirical, idiomatic) is evidence that a writer is referencing a private state.

The relation between adverbial and adjectival modifiers and the subset of subjective statements which are opinions, is the subject of many studies, and is still being actively investigated. Van Steenburgh (1987, p. 378) showed that, "[as with quickly] the presence of a scaling adverb is shown to be not property ascription... but by presupposed [pace] scale."

This characteristic of adverbs (and some adjectives) makes adverbial lexemes important markers in the identification of opinionated statements. Substantial body of research in both applied linguistics and opinion mining and sentiment analysis rely on this characteristic of adverbial lexemes (Xu et al., 2007a; Osman et al., 2007; Bethard et al., 2004; Pak & Paroubek, 2010).

A classification scale of attitudes was important to social scientists because, as Likert (1932, p. 7) described, "the number of attitudes which any given person possesses is almost infinite."

2.2.5 Propositions

The following propositions are this paper's response to the problem of access (Problem 2.1) and accuracy (Problem 2.2) discussed in Sections 2.1.1 and 2.2.4. They also summarize the aims of this research along the lines of Research Questions 2.1 and 2.2. Each is an affirmative statement to clarify arguments for or against what is being proposed.

Proposition 2.1. *Sentiment-type is an essential semantic element of opinion.*

The above proposition, Proposition 2.1, is justified because dimensionless sentiments do not exist in social psychology, social science, or any field excepting opinion mining: $+1$ and -1 are not sentiments.

Proposition 2.2. *A defined sentiment-type-scale is essential for opinion mining results to be meaningful.*

As with Proposition 2.1, Proposition 2.2 seems justified because scales of opinion measurement used in other disciplines, within a sentiment type, always have a scale, such as the ubiquitous Likert-scales.

Proposition 2.3. *There exists a canonical list of sentiment types.*

As discussed in Section 2.2.2, while the ways in which people can express emotions is limitless, the types of emotions seem limited. In the same way, while the number of ways in which people can hold a sentiment is limitless, the types of sentiment seem limited. Given the prospect of a limited set of sentiment types, Proposition 2.3 does not seem to be overreaching. The essence of sentiment is rooted in human emotion, which is a linguistically and culturally neutral construct. Therefore, it seems plausible that a canonical list of sentiment types for English language users would show some correspondence to users of other languages. The following experimental procedures were used to explore Research Questions 2.1 and 2.2. Propositions 2.1 to 2.3 are discussed in Section 2.5.

2.3 Methods

Two principal tasks are involved in exploring Research Questions 2.1 and 2.2 as discussed in Section 2.2.5. The first involves defining a taxonomy of sentiment types for English. The second is to examine whether or not those sentiment types are similarly represented in languages other than English.

2.3.1 Methodology

The experimental procedures used in this research are derived from the constructs and relationships presented in the theoretical construction of this study in Section 2.2, and summarized in Section 2.1.3. While the operationalization of a theoretical model can be represented in many different forms, logic notation seemed to be a concise, comprehensive, and convenient approach for this work. Also, logic notation enables a more specific form of cross-referencing between sections. The mnemonics of “P1” and “P2” refer to the proof associated with Task 1 and 2, respectively.

Task 1: Derive a Taxonomy of Sentiment Types

The following logic proof undergirds the methodological approach for deriving a set of sentiment types from social science scholarship on survey construction, as described in Research Question 2.1.

Task 1 Proof:

1. Let O be the canonical opinion frame.
2. Let E be the set of semantic elements which make up O.
3. Let s be the semantic element corresponding to the *private state*, *opinion**, or *way-of-thinking* (See Section 2.2.2) semantic element of E.
4. Let T be the universe of possible types of s.
5. Let $t \in T$.
6. Then $\forall s \ni t$.
7. Let $scale_t$ be a scale for measuring t.
8. Let t_o be a specific instance of t.
9. Let $scale_o$ be a scale for measuring type t_o.
10. Let o_{en} be a set of English-language instances of O which include t_o.
11. Let Q be the universe of possible semantically orthogonal survey questions.
12. Let $Q_{en} \subset Q$, in English-language.
13. Let $Q_{o_{en}} \subset Q_{en}$, designed to measure o_{en}.
14. Then, $Q_{o_{en}}$ will include questions which reference $scale_o$.
15. Then, $\hat{Q_{en}}$, a subset of Q_{en}, can be used to derive $\hat{T_{en}} \subset T$.

The goal, then, of this part of the research, is to find some $\hat{Q_{en}}$ which is representative of Q. This done, the next step is to derive a set of types, $\hat{T_{en}}$, which is representative of T_{en} or T, depending on the findings of the second portion of this research. The $\hat{Q_{en}}$ used, and specific heuristics applied to derive $\hat{T_{en}}$ are covered in Section 2.3.3 below.

Task 2: Measure Usage of Sentiment Types Across Languages

Assuming a $\hat{T_{en}}$ can be determined, the following logic proof undergirds the methodological approach for determining whether or not $\hat{T_{en}}$ is representative of T, as expressed in Research Question 2.2. This approach is similar to work done by Déjean, Gaussier & Sadat (2002) and also by De Melo & Weikum (2009).

Task 2 Proof:

1. Given $\hat{T_{en}}$ from Research Question 2.1.
2. Let C be a social media corpus of textual documents including remarks made by individuals in various languages, L.
3. Let $l \in L$, but not English.
4. Let r_x be the subset of the documents in C where the language used is x.
5. Let $t_{en} \in \hat{T_{en}}$.
6. Let $\tilde{t_{en}}$ be a set of adverbial or adjectival exemplars of t_{en}.
7. Assume that if any $\tilde{t_{en}}$ is present in r_{en} then t_{en} is present.
8. Let $\tilde{t_l}$ be a reliable translation $\tilde{t_{en}}$ for language l.
9. Assume that if any $\tilde{t_l}$ is present in r_l then $\tilde{t_{en}}$ is present, which implies t_{en} is present from *(P2.7)*.
10. Let $n_{t_{en}}$ be the counts of r_{en} in which t_{en} is present.
11. Let n_{t_l} be the count of r_l in which t_{en} is present.
12. Let $rank_{t_{en}}$ be the ordinal position of $n_{t_{en}}$ sorted in descending order.
13. Let $rank_{t_l}$ be the ordinal position of n_{t_l} sorted in descending order.

14. If $\forall l \, \forall t : \, rank_{t_l} \approx rank_{t_{en}}$ then conclude, $\forall l : \, T_l \approx \hat{T_{en}}$ *(language-portability)*.

The goal, then, of the second portion of the study is to evaluate whether or not $\hat{T_{en}}$ is represented similarly in non-English languages within social media content. The details on how this problem was analyzed are provided below in Section 2.3.3.

2.3.2 Hypotheses

Research Question 2.1 involved an investigative inquiry into the development of a taxonomy of sentiment types, so no testable hypotheses were defined. As shown in the proofs above and elsewhere, the outcomes from Research Question 2.1 are a direct input into Research Question 2.2. The operational model for the inquiry suggested by Research Question 2.2 (language-portability of a sentiment type taxonomy), however, is a testable null hypothesis.

Null Hypothesis 2.1. *When using a standard method of comparing sorted lists, the frequency of occurrence of sentiment types in English social media content will be significantly different from frequencies found in a corpus of non-English social media content.*

If the language-portability of the English taxonomy of sentiment types cannot be shown, then the value of that taxonomy is reduced significantly. The theoretical construction of concepts in Section 2.2.2 expresses some reliance on the universality of human emotion—and the subsequent universal effects expected in the creation of private states by humans expressing opinions.

Consistent with that logic, even if language-portability can be shown, it only provides a statement that the proposed sentiment type taxonomy is robust against the tested conditions. A substantial amount of further research would be required to affirm a positive finding in this study.

2.3.3 Procedures

The following experimental procedures are the practical expression of the methodology defined above for Task 1 and Task 2, designed to evaluate whether or not to reject Null Hypothesis 2.1 found in Section 2.3.2. An overview of the procedures which operationalize Task 1 and Task 2 is shown below in Figure 2.1.

Task 1 : Derive a Taxonomy of Sentiment Types for English

For the past 80 years, the survey has been the most reliable instrument for harvesting public sentiment. Therefore, we used survey design scholarship as the basis of our approach to developing a sentiment type taxonomy. Within survey design, Likert and Likert-like scales are a common standard for discretizing the way of thinking of a cognizer about a particular topic (Morgeson et al., 2006). Likert's original 5-point scale included strongly disagree, disagree, neither disagree nor agree, agree, and strongly agree.

Preparation. As shown in Figure 2.1, an inventory of 37 common types of Likert-like scales used in surveys to capture public sentiment (Vagias, 2006). However, because Vagias (2006) contains some duplication and potentially overlapping concepts the list of 37 scales needed to be condensed. An example of duplication is the five scales used for frequency. An example of overlapping concepts is the use of both problem and difficulty. Table 2.1 shows the original Vagias (2006) inventory and subsequent redactions.

After this initial redaction, there were 26 lexically unique sentiment scales.

WordNET hypernym semantic distances were calculated for each scale to meet the semantic orthogonality requirement of P1.11. The frequency of occurrence was also calculated for each scale, to identify the principal representation for those

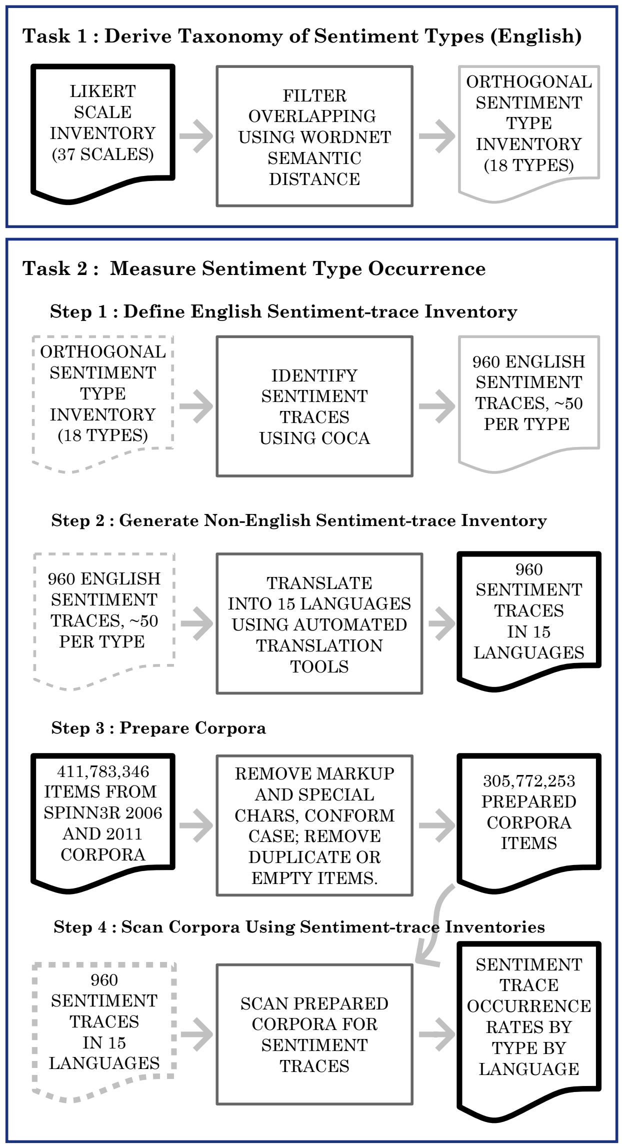


Figure 2.1: The experimental procedures for Task 1 and Task 2.

Table 2.1: Summary of survey scales.

	Vagias (2006) Likert-type Scales (37)	Measurement	Lexically Unique (26)		WordNet Distinct (20)	COCA Search Results Query	Forms	Occurs
1	level of acceptability	acceptability	acceptability		(dist 2) quality			
2	level of appropriateness	appropriateness	appropriateness		(dist 2) quality			
3	level of importance	importance	importance	**1**	importance	[r*] *important*	77	67038
4	level of agreement	agreement	agreement	**2**	agreement	[r*] *agree	48	4287
5	knowledge of action	trueness	trueness	**3**	trueness	[r*] true | false	65	8011
6	reflect me?	trueness	(lexical duplicate)					
7	my beliefs	trueness	(lexical duplicate)					
8	priority	priority	priority	**4**	priority	[j*] priority	16	3130
9	level of concern	concern	concern	**5**	concern	[r*] *concerned	51	8664
10	priority level	priority	(lexical duplicate)					
11	priority problem	problem	problem		(dist 1) difficulty			
12	affect on X	effect	effect	**6**	effect	[r*] affect	26	1972
13	level of consideration	consideration	consideration	**7**	consideration	[r*] consider	25	2839
14	level of support/opposition	support	support	**8**	support	[r*] support | oppose	42	2403
15	level of probability	probability	probability		(dist 1) likelihood			
16	level of agreement	agreement	(lexical duplicate)					
17	level of desirability	desirability	desirability		(dist 1) quality			
18	level of participation	pariticipation	pariticipation		(dist 2) difficulty			
19	frequency 5 point	frequency	frequency	**9**	frequency	[r*] never | rarely | occassionally | sometimes | frequenty | usually	76	8736
20	frequency 5 point (2)	frequency	(lexical duplicate)					
21	frequency of use	frequency	(lexical duplicate)					
22	frequency 7 point	frequency	(lexical duplicate)					
23	amount of use	frequency	(lexical duplicate)					
24	level of familiarity	familiarity	familiarity	**10**	familiarity	[r*] *familiar	43	4545
25	level of awareness	awareness	awareness	**11**	awareness	[r*] *aware	62	6046
26	level of difficulty	difficulty	difficulty	**12**	difficulty	[r*] difficult*	55	26595
27	likelihood	likelihood	likelihood	**13**	likelihood	[r*] *likely	46	34671
28	level of detraction	detraction	detraction		detraction	[r*] detract	none	
29	good/bad	positiveness	positiveness	**14**	positiveness	[r*] positive | negative	67	6625
30	barriers	barrier	barrier		barrier	[r*] bar	none	
31	level of satisfaction 5 point	satisfaction	satisfaction	**15**	satisfaction	[r*] *satisf*	53	2850
32	level of satisfaction 5 point (2)	satisfaction	(lexical duplicate)					
33	level of satisfaction 7 point	satisfaction	(lexical duplicate)					
34	level of quality 5 point	quality	quality	**16**	quality	[r*] good | poor | bad | excellent	104	75262
35	comparison of two products	goodness (relative)	(lexical duplicate)					
36	level of responsibility	responsibility	responsibility	**17**	responsibility	[r*] responsible	43	4294
37	level of influence	influence	influence	**18**	influence	[r*] *influen*	61	5201

terms whose semantic distances are small. The Corpus of American Contemporary English (COCA) (Davies, 2009) was used to calculate the frequency of occurrence.

Analysis. Those scales whose semantic distance was <3 were collapsed into the term with the highest frequency of occurrence in COCA, to reduce the list to orthogonal terms. This approach to ensuring semantic orthogonality is modeled in part after (De Melo & Weikum, 2009), who used WordNET semantic distances and automated translation to extend the is-a and has-a definitions to other languages.

Of the 37 original Likert-like scales, 26 were found to be lexically unique, and of those, 20 were found to represent semantically distinct sentiment types. For example, “appropriateness” and “acceptability” were removed, as each had a semantic distance of 2 to the sentiment type: “quality” whose frequency of occurrence was higher in COCA. Two sentiment types, “barrier” and “detraction,” were removed for paucity because each had fewer than 20 occurrences in COCA. Table 2.1 shows the original list of 37 sentiment types from Vagias (2006), the search clauses used to measure frequency of occurrence in COCA, and the final list of 18 sentiment types. The number of COCA occurrences returned for each search and the number of unique sentiment traces found are also shown. These 18 sentiment types correspond to the $\hat{T_{en}}$ taxonomy of sentiment types from Proof P1.15, above.

Task 2 : Measure Usage of Sentiment Types Across Languages

As shown in Proof P2.1-2, Task 2 uses $\hat{T_{en}}$ to measure the frequency of sentiment types t within some corpora C. The complete 4-step process for Task 2 is shown graphically in Figure 2.1. Where there is some direct correspondence to Proofs P1 and P2, those cross-references are provided.

Step 1 (P2.5-6). Because the sentiment types in $\hat{T_{en}}$ (i.e., “quality”, “importance”) are not themselves adverbial or adjectival scaling modifiers, an inventory of corresponding modifiers is needed. These scaling modifiers were called sentiment traces in Section 2.2.2. The presence of a sentiment trace, therefore, is strong indicator indication of sentiment. It is difficult for a person to use those lexical elements without reliance on a private state (Breck et al., 2007; Tang et al., 2008) for its meaning.

The Corpus of American Contemporary English (COCA) (Davies, 2009) provides a robust query engine and was used to identify the most commonly used sentiment traces for each of the 18 remaining sentiment types. Controlling for ambiguity between sentiment traces was beyond the scope of this inquiry. An example of ambiguity is “national priority.” Uncommon grammatical constructions are also ignored, such as “about positive” or “here agree.” Despite these challenges, a manual inspection of search results did not reveal any undue influence of such cases. For example, Table 2.3 shows the 48 unique sentiment traces returned by COCA for the sentiment type “agreement” using the search string “[r*] *agree.” Examples 2.3 to 2.5 show samples of the COCA corpus occurrences for “agreement,” with the sentiment traces underlined.

Example 2.3. “...I would say one other thing about Cheney though, and I <u>certainly agree</u> with Charles when he said Cheney showed how a vice president can have influence.”

Example 2.4. “...I think everybody in this room would <u>probably agree</u> with that. Anyway, let’ s see...”

Example 2.5. “...complaints on one side or the other is not the best measure. I <u>totally agree</u> with you that - I think perhaps a better measure of how people perceive us...”

Table 2.2: COCA search results by sentiment type ([r*] = adverb, j* = adjective)

Sentiment Type	COCA Search	Lines	Traces
quality	[r*] good|poor|bad|excellent	75262	104
importance	[r*] *important*	67038	77
likelihood	[r*] *likely	34671	46
difficulty	[r*] difficult*	26595	55
frequency	[r*] never|rarely| occassion ally|sometimes|frequenty| usually	8736	76
concern	[r*] *concerned	8664	51
trueness	[r*] true|false	8011	65
positiveness	[r*] positive|negative	6625	67
awareness	[r*] *aware	6046	62
influence	[r*] *influen*	5201	61
familiarity	[r*] *familiar	4545	43
responsibility	[r*] responsible	4294	43
agreement	[r*] *agree	4287	48
priority	[j*] priority	3130	16
satisfaction	[r*] *satisf*	2850	53
consideration	[r*] consider	2839	25
support	[r*] support|oppose	2403	42
effect	[r*] affect	1972	26

Table 2.3: Sentiment traces returned by COCA for sentiment type "agreement"

Sentiment Trace		Sentiment Trace	
1	strongly disagree	25	just agree
2	strongly agree	26	completely disagree
3	also agree	27	ever agree
4	always agree	28	really disagree
5	totally agree	29	only agree
6	certainly agree	30	wholeheartedly agree
7	generally agree	31	still agree
8	probably agree	32	finally agree
9	completely agree	33	also disagree
10	absolutely agree	34	much agree
11	now agree	35	often disagree
12	even agree	36	basically agree
13	necessarily agree	37	neither agree
14	never agree	38	readily agree
15	respectfully disagree	39	often agree
16	totally disagree	40	largely agree
17	fully agree	41	entirely agree
18	both agree	42	still disagree
19	actually agree	43	vehemently disagree
20	just disagree	44	usually agree
21	quite agree	45	here agree
22	really agree	46	agree
23	definitely agree	47	disagree
24	of agree	48	agree/disagree

Table 2.4: Translations of the sentiment trace "strongly disagree"

English	**Strongly Disagree**
Arabic	لا أوافق بشدة
Chinese	強烈反對
Chinese-Simp	強烈反对
German	stimme überhaupt nicht zu
Spanish	muy en desacuerdo
French	fortement en désaccord
Italian	molto in disaccordo
Japanese	強く反対
Korean	강력 반대
Dutch	zeer oneens
Portuguese	discordo
Russian	решительно не согласен
Swedish	starkt emot
Ukrainian	рішуче не згоден

The number of English sentiment traces for each of the sentiment types is shown in Table 2.2. The final inventory contained 960 English sentiment traces, which are the $\tilde{t_{en}}$ referenced in Proof P2.6.

Step 2 (P2.8). In this step the English sentiment traces, $\tilde{t_{en}}$, were translated into the predominant languages represented in the corpus. These languages included: Russian, Japanese, Chinese, Chinese- Simplified, Spanish, German, French, Italian, Portuguese, Dutch, Swedish, Ukranian, Arabic, and Korean.

Translation was done using an automated translation tool provided by Google. While the Google translation service is fairly new, it has been successfully used in research for some of the target languages (Wan, 2008; Kursten et al., 2008). An informal reasonableness check was done with native speakers on some of the translations. The consensus amongst these reviewers was that the automated tool provided a plausible translation. Each pointed out the substantial influence that context has on an interpretation. The Swedish reviewer commented that, "Some of these phrases you would never hear a native speaker say or write, but the connotation is there." Table 2.4 shows the translation into 14 non-English languages of one of the 960 sentiment traces, "strongly disagree."

Some consideration was given to performing a comprehensive human review of the automated translations. However, it was felt that human intervention might undermine experimental repeatability. Human modification may also limit the lessons learned regarding the benefits and limitations of automated translation. Therefore, the Google translations of the 960 sentiment traces were used unaltered. At the conclusion of Step 2, the complete inventories of sentiment traces for non-English languages were derived. These inventories are referred to as $\tilde{t}_l$ in Proof P2.8.

Step 3. In this step, the corpora are prepared for analysis. The two corpora chosen for this research come from the same source but are three years apart in their origin (Burton, Java & Soboroff, 2009; Burton, Kasch & Soboroff, 2011). The 2009 ICWSM dataset is from 2008 while the ICWSM 2011 dataset is from 2011.

The availability of this type of longitudinal corpora presents a significant opportunity to examine the robustness of the findings of this research across a formative period in the development of social media. The Spinn3r Internet Feed corpora (Burton et al., 2009, 2011) used for this research contains 411,783,346 documents provided in XML formatted files in languages relevant to this study. Other languages were available, but only the top 15 languages (including English) were considered here because they represent 97.4% of the documents. Romanian documents were excluded in deference to Korean documents, despite Romanian documents being present in slightly higher concentrations than Korean. This exchange was done because of the high number of CJK (Chinese, Japanese, and Korean) documents in the corpora were expected to include Korean language documents. The total corpora size for all languages was approximately 300GB, compressed.

Metadata about each document was included in the corpora: language, source website or URL, publication type (or, register), and post (the authored content.) The register, language, and post datums were the ones relevant to this research. The unique values for register in the corpora were: CLASSIFIED, FORUM, MAINSTREAMNEWS, MEMETRACKER, REVIEW, SOCIAL-MEDIA, AND WEBLOG.

The 2008 corpus included an UNKNOWN register type. This register accounted for 16% of that corpus. However, 93% of these were from livejournal.com, a well-known blog site. Therefore, corpus documents whose register was marked as UNKNOWN but which were from livejournal.com were reclassed to the WEBLOG register.

To prepare each document for scanning, all HTML and XML tags, duplicative whitespace, and special characters were removed. Examples 2.6 and 2.7 show an Arabic document before and after preparation. Glosses for non-English documents are not provided for the examples because they are not being presented for their semantic content but their lexical and symbolic content.

Example 2.6. Arabic document before preparation.

<div style="direction:rtl;textalign: right "> لی یوبیزیت
که ب ش يً◌ت یآ ها ردای ارگ ک ه یب عىد س ه کدهً◌و ی چهار در » یجارخا »یهجوىعه
</div>. رود رو ههرهاه نً◌ه اهروز از ،سوت یق م سه

Example 2.7. Arabic document after preparation.

یهجوىعه لی یوبیزیت
ها ردای ارگ ک ه یب عىد س ه کدهً◌و ی چهار در یجارخا
رورود ههرهاه نً◌ه اهروز از ،سوت یق م سه که ب ش يً◌ت یآ

Table 2.5: Corpora document counts by author language

	Documents	Duplicates	<3 Words	Total Included/%	
Totals	411,783,346	76,621,887	29,389,206	305,772,253	74.3%
ARABIC	1,463,519	269,388	24,094	1,170,037	79.9%
CJK	11,285,988	3,519,658	3,906	7,762,424	68.8%
DUTCH	6,022,402	2,770,584	14,199	3,237,619	53.8%
ENGLISH	**250,552,022**	**46,959,977**	**3,713,179**	**199,878,866**	**79.8%**
FRENCH	6,451,492	2,570,213	47,767	3,833,512	59.4%
GERMAN	7,549,237	3,424,837	20,086	4,104,314	54.4%
ITALIAN	4,093,903	1,425,916	94,629	2,573,358	62.9%
JAPANESE	8,084,118	1,077,693	1,055	7,005,370	86.7%
KOREAN	620,930	190,286	26,373	404,271	65.1%
PORTUGUESE	2,758,721	362,804	8,451	2,387,466	86.5%
RUSSIAN	5,586,044	438,340	8,866	5,138,838	92.0%
SPANISH	6,928,110	2,115,378	35,559	4,777,173	69.0%
SWEDISH	2,418,085	1,016,887	11,254	1,389,944	57.5%
UKRAINIAN	442,081	33,098	530	408,453	92.4%
UNKNOWN	97,526,694	10,446,828	25,379,258	61,700,608	63.3%

To further prepare the corpus for analysis, duplicate and lexically poor (<3 words) documents were removed. We found that the level of duplication was much higher in the 2011 corpus (20.0%) than the 2008 corpus (10.7%). Conversely, we found that the number of lexically poor documents was much higher in the 2008 corpus (12.4%) than the 2011 corpus (6.2%). Redactions reduced the overall corpora document count by 26.7%. At this point in our process, we have a single corpus which has 305,772,253 documents. The document counts and redactions are shown in Table 2.5.

Step 4 (P2.10-11). In this step, the 960 sentiment traces for each of the 15 languages is used to build a library of regular expressions capable. This library of regular expressions was used to detect an occurrence of each of the 18 sentiment types per Proof P2.7. Care had to be taken regarding the significance of whitespace, as Chinese, Japanese, and Chinese-simplified do not necessarily use spaces between words. The Dutch and Arabic regular expression contain emphatic space

Table 2.6: Regular expression samples from Chinese, Dutch, and Arabic

Language	Sentiment Trace Regex
CHINESE AGREEMENT	**強烈反對|非常同意|也同意|總是同意|完全同意|當然同意|普遍認同|可能同意|完全同意|絕對同意|現在同意|即使同意|一定同意|絕對不同意|恭敬地不同意|完全不同意|完全同意|雙方同意|其實也同意|只是不同意|十分同意|真的同意|絕對同意|的同意|剛剛同意|完全不同意|怎會同意|真的不同意|只同意|完全同意|還同意|終於同意|也不贊成|認同|往往不同意|基本上同意|不同意|輕易同意|常同意|基本上同意|完全同意|仍然不同意|強烈不同意|通常同意|在這裡同意|同意|不同意|同意[ ]不同意**
DUTCH PRIORITY	topprioriteit | hoge[ ]prioriteit | hoogste[ ]prioriteit | een[ ]hogere[ ]prioriteit | lage[ ]prioriteit | nationale[ ]prioriteit | bestuurlijke[ ]prioriteit | een[ ]lagere[ ]prioriteit | belangrijke[ ]prioriteit | belangrijke[ ]prioriteit | belangrijkste[ ]prioriteit | meer[ ]prioriteit | dringende[ ]prioriteit | wetgevende[ ]prioriteit | de[ ]laagste[ ]prioriteit | prioriteit
ARABIC CONCERN	]أأقلل|ججددا[ ]ققللققة|ججددا[ ]ققللقق|ااههتتمماامم[ ]أأككثثرر]ممنن[ ]أأككثثرر[ ]االلممععننييةة|ححققاا[ ]ققللقق|أأييضضاا[ ]االلققللقق|ااههتتمماامم[ ]االلممققاامم[ ]ففيي[ ]االلممععننييةة|خخااصص[ ]ببققللقق|ععممييقق[ ]ققللقق|غغييررههاا[ ]ققللقق|ممتتززاايدد[ ]ققللقق|ججددا[ ]ققللقق|االلممععننييةة[ ]ككممااققللييللاا[ ]ققللقق|االلأأوولل[ ]ببششككلل[ ]االلققللقق|االلممععننييةة[ ]ككييفف|ققللققةة[ ]تتززاالل[ ]لل|ووااضضحح[ ]ببششككلل[ ]ققللقق|لللغغاايةة[ ]ققللققةة|االلممععننييةة[ ]ففققطط|دداائممةة[ ]ققللقق|خخااصص[ ]االلأأططرراافف[ ]ججممييعع|ححققيييقي[ ]ققللقق|ففققطط[ ]االلممععننييةة|ججددا[ ]ببننففسس|ججددا[ ]ققللققةة|ششدديدد[ ]ببققللقق|االلققللقق[ ]ممنن[ ]ببششييء|االلممععننييةة[ ]ييتتععللقق|ببااللغغ[ ]ببققللقق[ ]تتششععرر|االلققللقق[ ]ممنن[ ]االلققددرر[ ]ممععننييييننأأسساسساا[ ]تتتتععللقق|ووااضضحح[ ]ببششككلل[ ]االلممععننييةة|ممععظظممههاا[ ]ووإإذذ|االلآآنن[ ]االلممععننييةة|ممببااششررةة[ ]االلششييء[ ]ببععضض[ ]ققللقق|ككثثييرراا[ ]ححدد[ ]إإللىى[ ]االلممععننييةة|ببااللببييئةة[ ]االلممههتتممييينن|االلققللقق[ ]ببااللغغ[ ]ييسسااوورره[ ]االلووااضضحح[ ]ووممنن|االلممععننييةة[ ]االلأأققلل|أأططوولل[ ]االلممععننييةة|ححققاا[ ]ققللقق|ككبييرر[ ]ققللقق|تتههمم[ ]االلتتيي|ححتتىى[ ]االلممععننييةة|االلممععننييةة[ ]أأببددا|االلققللقق[ ]أأنن[ققللقق[ ]غغييرر|ققللقق|االلممععننييةة[ ]بببسساططةة|ااجتتمماععيياا[

markers between words because these languages include spaces and syntactic word boundaries.

Word boundaries are not used in the Japanese or Chinese regular expressions because the written language itself does not include spaces or other obvious syntactical boundaries. Table 2.6 shows samples of a portion of the generated regular expressions for Chinese, Dutch, and Arabic. The pipe-character "|" shown below represents a logical or. The Dutch regular expression for priority is shown in Table 2.6. It contains both word boundary markers and emphatic space markers between words because the Dutch language includes spaces and syntactic word boundaries.

As shown in Table 2.5, approximately 97M documents were marked as language UNKNOWN. The documents marked as UNKNOWN language were scanned for sentiment traces in all 15 languages as a pre-test. The resulting hit rate yielded a very low 8.2% detection rate—indicating that the UNKNOWN languages are likely outside of the set of 15 languages selected for this study. By way of comparison, corpus documents marked as CJK were scanned for sentiment traces in Chinese, Chinese-simplified, Japanese, and Korean, yielding a 63.2% detection rate. Because the linguistic character of the UNKNOWN language documents was so dissimilar from the documents in other known languages, the UNKNOWN documents were left out of the final scan for sentiment traces. After a substantial amount of preparation, the corpora documents are ready for scanning, and the scanners for all 15 languages are available.

Scanning 300M documents for occurrences of 960 sentiment traces in 15 languages is a computationally intensive task. Scanning engines were written in C++, Python, Perl, and Node.js. The final version, written in Node.js, performed 10% slower than a C++ partial implementation but was very stable and easy to develop. The other implementations were not comparatively efficient. Figure 2.2 shows the linkage between the social science scholarship in Task 1 and the opinion mining and sentiment analysis canonical model of opinion which is augmented by the taxonomy of sentiment types derived from Task 2. The elements which are grayed out in Figure 2.2 are those elements in the canonical model of opinion, but which are not the subject of this research directly. The results derived from this process are discussed next, in Section 2.4.

2.3.4 Disclosures

No human subjects were used in the execution of this study. This research was otherwise conducted in accordance with the guidelines published by Iowa State University Institutional Review Board regarding the protection of human participants in the Investigator Handbook.

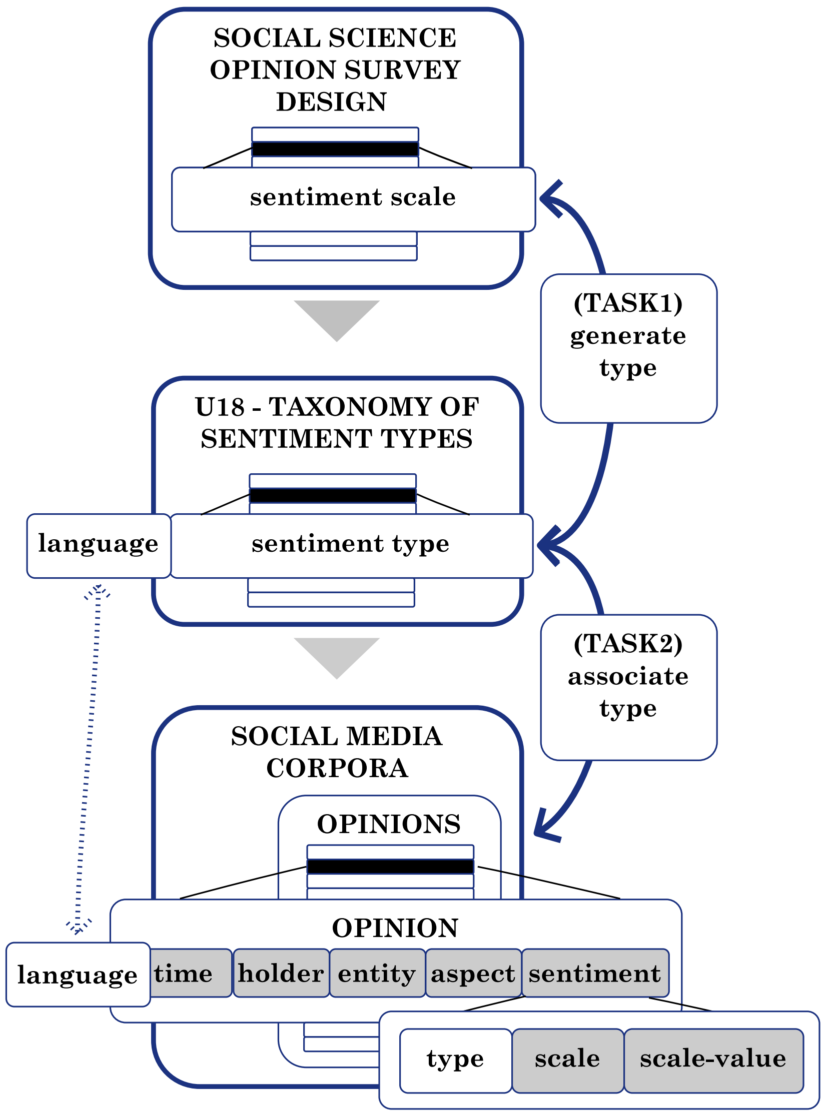


Figure 2.2: Operational view of the linkage between social science sentiment scale inventory and the canonical model of opinion commonly used in opinion mining research

2.4 Results

Before the results for evaluating Null Hypothesis 2.1 are discussed, two findings emerged in the course of this research which shed light on the veracity of the data and procedures.

2.4.1 Sentiment Trace Density by Register

The documents in the corpora used in this research are classified by register, a formalism for what could loosely be called distinct combinations of genre and channel. Two registers in this corpora are of particular importance to this research, namely WEBLOG and SOCIAL-MEDIA, as the focus of this research is sentiment expression in social media.

A question arose about whether or not it was appropriate to combine the two registers to develop a single taxonomy of sentiment types for social media. Social media is often described as including both WEBLOG and shorter more transient messaging platforms such as Twitter, which is the primary source for the SOCIAL-MEDIA documents. For example, Huang (2013, p. 14) characterizes social media as, "social media data, e.g., comments, blog articles, or tweets." Qualitatively, SOCIAL-MEDIA represents a terse quasi-prose while WEBLOG entries tend to be longer expositions.

An examination of the sentiment trace densities from both SOCIALMEDIA and WEBLOG documents might say something interesting about how social media users express themselves in the two registers. While not in the scope of the primary thrust of this research, the findings were interesting enough to report, here.

Word counts for all documents in the corpora across all languages and registers were calculated. Word counts for non-pictographic languages were determined by

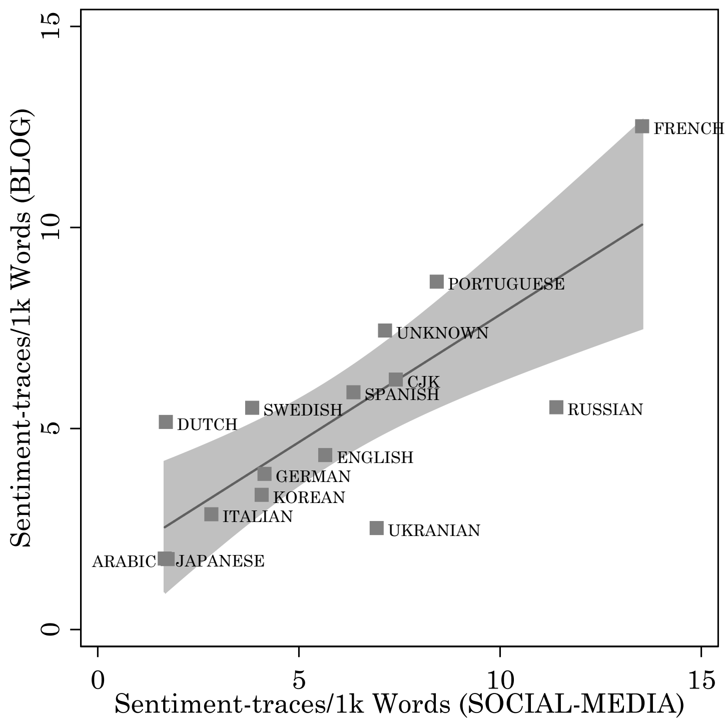


Figure 2.3: Comparative values for sentiment trace density by language between SOCIAL-MEDIA and WEBLOG registers

counting the number of spaces in a conformed version of the document and then adding 1. For Chinese and Japanese, the conventional formulas of 2.4 characters per word for Chinese (Marcus et al., 1994) and 2.1 characters per word for Japanese (Green, 1999) were used.

As shown in Figure 2.3, while there is a slight shift toward SOCIAL-MEDIA having a higher density of sentiment traces than WEBLOG, the differences are not significant. A paired t-test of the difference between sentiment trace densities for WEBLOG and those of SOCIAL-MEDIA by language did not show a statistically significant average difference ($p<0.3$).

This finding is important for two reasons. First, it represents the first comparative look at sentiment density across languages. Secondly, it demonstrates that authors of WEBLOG and SOCIAL-MEDIA documents use sentiment traces at statistically similar rates. This finding enables a level of generalization about the nature of sentiment expression between the two registers.

Moreover, this finding lends some notional support for the proposition that social media users of different languages express differing relative amounts of sentiment-laden content in social media.

2.4.2 Sentiment Trace Occurrence Rates for English

While prior research on sentiment trace density by language does not exist, a few other researchers have measured subjective content densities for English. Therefore, it made sense to use this related research as a quality check against the findings presented in this research.

As shown in Figure 2.4, it was found that the overall sentiment trace occurrence rate was 512.1 per 1,000 documents. Traces of the quality sentiment type were the most frequently occurring at 212.7 per 1,000 documents. This result is consistent with Macdonald et al. (2007). Macdonald described a social media corpus contain-

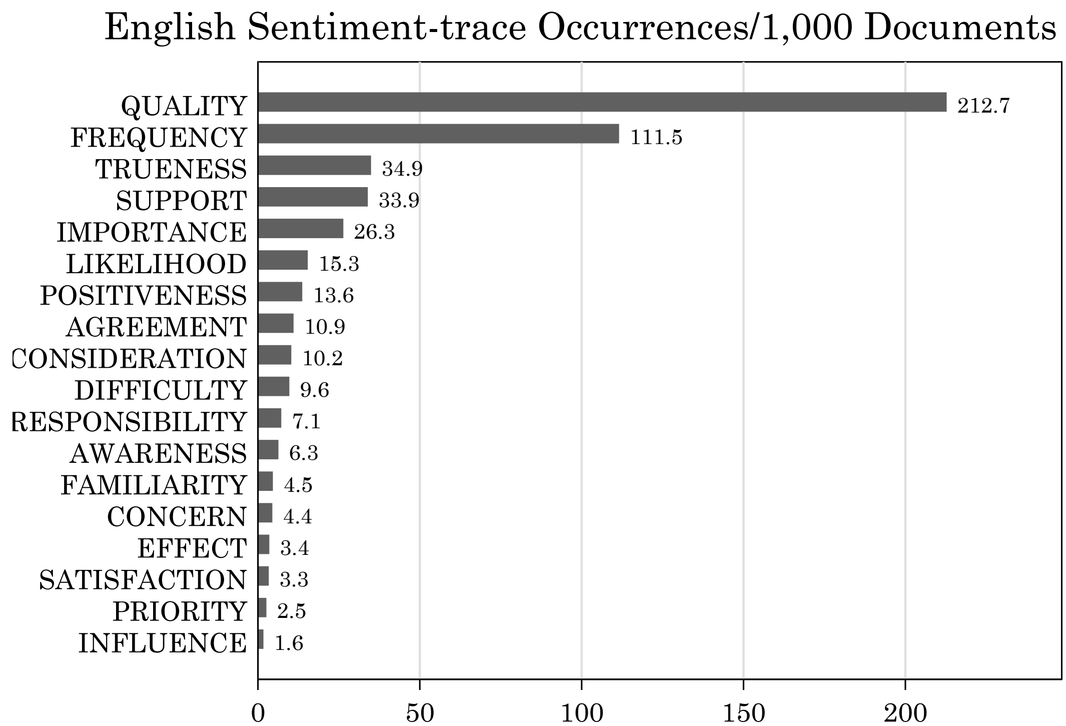


Figure 2.4: English sentiment trace occurrences by sentiment type

ing 579.7 opinions per 1,000 posts. Fu & Wang (2010) found that 62% of Chinese documents in a corpus contained opinionated content. Though these figures do not inform Research Questions 2.1 and 2.2 directly, they do support Proof P2.6, which asserts that $\tilde{t_{en}}$ is a set of exemplars indicative of subjective statements.

2.4.3 Sentiment Type Density By Language

To restate the hypothesis in simpler terms, do users of social media who write using different languages express themselves using similar types of sentiment at similar frequencies? For example, do Arabic and Swedish social media users discuss "quality", "agreement", or "support", in similar proportions to those who use English?

Rank order correlation values were calculated between languages and across the different corpora. The frequency of sentiment type occurrence was used as the

Table 2.7: Spearman rank order correlation significance values for each language

Language	Combined	2011 Corpus	2008 Corpus
NON-ENGLISH	p<.001***	p<.001***	p<.001***
ARABIC	p<.01**	p<.01**	**p<.1***
CN	p<.001***	p<.001***	p<.001***
CN-SIMP	p<.01**	p<.01**	p<.01**
GERMAN	p<.001***	p<.001***	p<.001***
SPANISH	p<.001***	p<.001***	p<.001***
FRENCH	p<.001***	p<.001***	p<.01**
ITALIAN	p<.001***	p<.001***	p<.001***
JAPANESE	p<.01**	p<.01**	p<.01**
KOREAN	p<.001***	p<.001***	p<.001***
DUTCH	p<.001***	p<.001***	p<.001***
PORTUGUESE	p<.001***	p<.001***	p<.001***
RUSSIAN	p<.001***	p<.01**	p<.01**
SWEDISH	p<.01**	p<.01**	p<.001***
UKRAINIAN	p<.01**	p<.01**	p<.01**

dependent variable. These values correspond to the measures $rank_{t_{en}}$ and $rank_{t_l}$ from Proof P2.12—which enables the evaluation of Null Hypothesis 2.1.

The rank order for each sentiment type for each language is shown in the heatmap in Figure 2.5. The sentiment trace occurrences were aggregated by language and type of sentiment, and ranked from the highest density (1) to the lowest (18). The leftmost column is the rank order of the sentiment type densities for English. The second column is a combined ranking of all non-English languages, and the remaining columns show the rankings across the other 14 languages. Figure 2.6 shows a scatter-plot of the sentiment type density rankings for all languages when compared to English as a baseline.

Spearman's rank-order correlation test was used to evaluate the rank order correlation between sets of U18 sentiment type densities. The p-values for these correlation relationships are presented in Table 2.7.

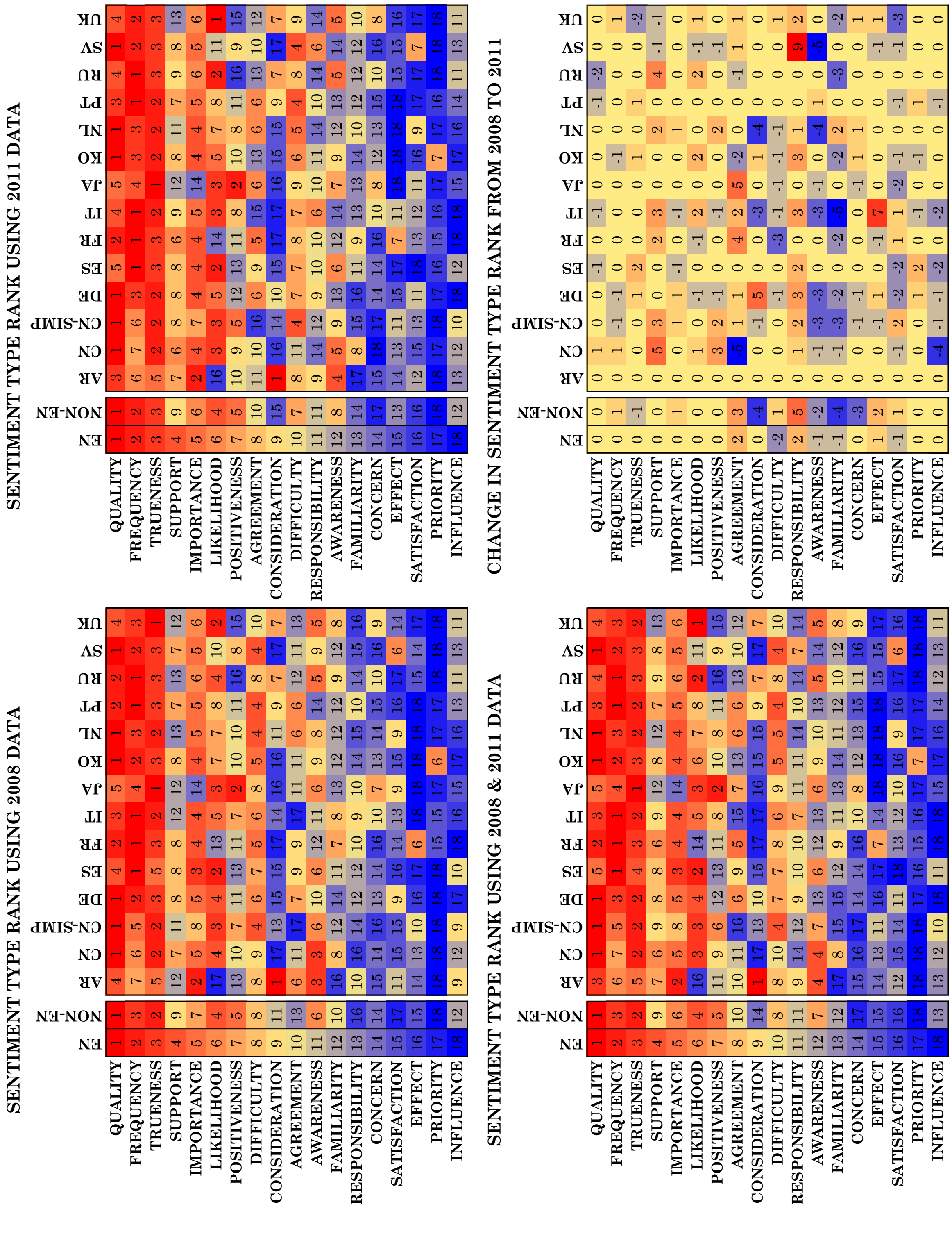


Figure 2.5: Heatmap of sentiment type density rankings by language and corpora

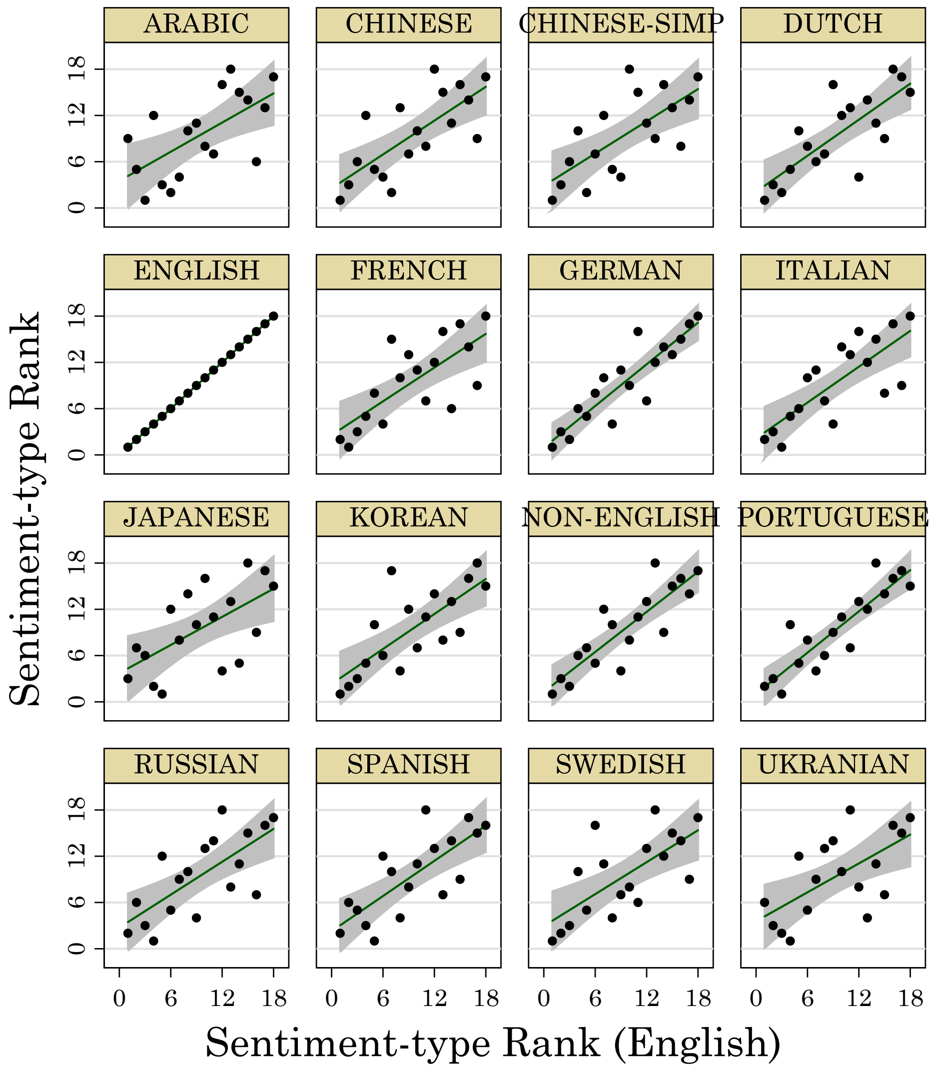


Figure 2.6: Scatter-plot of sentiment type density rankings by language

When English sentiment type rankings are compared with the rankings of other languages, the rank-order correlations appear to be very strong (most are p<0.001***). Additionally, when English is compared to the consolidated scores of Non-English languages, the relationship also appears significant (p<0.001***). These results are consistent across both the 2011 and 2008 corpora. The Arabic correlation in the 2008 corpus is the lowest. Interestingly, when compared to ENGLISH-COCA, the source for the sentiment traces, the results were at the lowest statistically significant level (p<0.1*).

The rejection of Null Hypothesis 2.1 is consistent with Proof P2.14 precondition of $\forall l\ \forall t\ :\ rank_{t_l} \approx rank_{t_{en}}$. This finding also provides support for the claim of this research that $\forall l : T_l \approx \hat{T_{en}}$—which can be understood to mean that the social media sentiment type taxonomy developed through this research is portable across languages.

2.5 Conclusions

Sentiment and opinion expression is a substantial portion of social media content, and a unique window into "what's on people's minds." With the global proliferation of devices capable of participating in social media, the diversity and depth of concerns expressed are extensive. Opinion mining and sentiment analysis research has lagged in the depth and breadth of reliable theoretical models which enable systematic study of this important phenomena.

The innovative aspects of this research include the integration of linguistic analysis, social science research on survey construction, and corpus linguistics. Some criticism of conventional methods of opinion mining and sentiment analysis research were discussed in Section 2.2.3. The goal of this research was to address those concerns and expand the dimensions of analysis for opinion mining and sentiment analysis researchers. The approach was to develop a taxonomy of sentiment types which is portable across languages. This goal was largely realized. However, some cautions are discussed below.

The guiding null hypothesis (Null Hypothesis 2.1) was summarily rejected, allowing a cautious first step toward concluding that a cross-language taxonomy of sentiment types does exist. In practical terms, this cautious first step postulates that populations of social media users who write using difference languages, do seem to express similar types of sentiment. Arabic speakers discuss "quality" of things about as frequently as Swedish, Spanish, English, or French speakers. For this to be useful, the corpus used in this research needs to be representative of the larger population of social media content—not a trivial assumption.

The question of generalizability of these findings to content outside the corpora used in this research is a natural one. With the rate of change in the composition of social media, any claim which reaches too far into the future is justifiably challenged. However, in view of the scale of the corpus (400M+ documents), the diversity of languages represented (15+), and the longitudinal character of the methodology (corpora separated by 3 years), it seems reasonable to attach enough validity to these findings to dig further.

Limitations

Extra care is taken with the discussion on the limitations of this research. This study is an attempt to define new constructs, integrate disparate disciplines, and generalize to a dynamic area of human discourse. In short, a lot is new, here. This research may spawn more questions than it answers and certainly includes weaknesses and limitations. Some of these are noted below.

The Proofs 2.1-2 will be used to anchor the criticism, since the logical construction affords a structural view of the approach used in this research.

Proof P1.11 states, "Let Q be the universe of possible semantically orthogonal survey questions." This study does not meet the test of that statement. The Vagias (2006) compilation of sentiment scales was the only comprehensive source of sentiment scales identified through the literature search. It is possible that other more complete inventories exist, and as such may provide a richer set of core terms from which to proceed.

Proof P1.15 concludes, "Then, $\hat{Q_{en}}$, a subset of Q_{en}, can be used to derive $\hat{T_{en}} \subset T$." However, no statement of completeness is included in the proof—only that $\hat{T_{en}}$ represents a valid subset. The findings of this research certainly support that claim. However, it would be a mistake to extrapolate from this claim that $\hat{T_{en}}$ somehow is a comprehensive taxonomy, relative to T ("the universe of possible types of s", Proof P1.3).

Another caveat to Proof P1.11 is that an extensive semantic analysis of each of the final 18 sentiment types would almost certainly produce a clarified typology with more, fewer, or different types. The automated method used for disambiguation in this research lacks the rigor found in more serious scholarship on disambiguation, such as Tsang & Stevenson (2010). In Proof P2.6 ("Let $\tilde{t_{en}}$ be a set of adverbial or adjectival exemplars of t_{en}."), no effort was made to control for ambiguity, idiom, sarcasm, nested-expressions, negation, or garden-path effects.

The inadequate treatment of each of these artifacts in this research certainly adds a level of noise to its findings. It is not beyond reason to suggest that they might overwhelm the results obtained. However, the innovative nature of this inquiry demands the taking of some risks, and future related research will no doubt improve our understanding beyond this initial attempt.

Also, it is also important to note that Davies (2009) corpus used to identify the sentiment traces is a media text corpus, not a social media corpus. While the English usage may be similar to that found in social media, it may not. No effort was made to examine whether or not the use of subjective language in COCA is different than the use of subjective language in social media. This potential problem has credence considering the finding that the rank order correlation between English sentiment types in the social media corpus and COCA were at the lowest statistically significant level (p<0.1*). If social media corpus indexed and searchable like COCA was available, that would have been used.

Proof P2.4 states, "Let $\tilde{t_l}$ be a reliable translation $\tilde{t_{en}}$ for language l." This assertion presents some cause for concern, as the automated translation of 960 English sentiment traces into 14 different languages will certainly introduce some noise into the findings of this research. While a manual review of some of the translations was done to check for reasonableness, a comprehensive evaluation was not done.

In Proof P2.1, the corpora, C, selected for counting sentiment trace frequency for this research spans a formative period in the proliferation of social media. As discussed in Section 2.4.1, the Burton et al. (2009) corpus included WEBLOG content, while the Burton et al. (2011) corpus included a SOCIAL-MEDIA register. Social media as a whole is a larger phenomena that includes micro-blogs (Twitter) and comment boards not referenced in the Burton et al. (2009) corpora.

While the literature review did not bring to light any relevant linguistic difference exists across these forms, these findings support the claim that there is a level of commonality between these registers. It also seems likely that some differences in the linguistic construction of opinion exists across languages. Those differences could influence these findings.

In Proof P2.1, the corpora, C, is not described as undergoing any conforming or normalizing process. However, to simplify the definition of the library of regular expressions used in this research, all special characters, numbers, and punctuation were replaced with whitespace.

No effort was made to deal adequately with hyphens. No effort was made to maintain references to phrasal terminals such as the period, semi-colon, colon, or comma.

As shown below in Example 2.8, the removal of symbols removes some aspects of meaning intended by the authors of the social media text. This alteration can lead to the false flagging of some sentiment trace occurrences. In the example, the syntactic separation between the word “really” and the word “difficult” is removed. This removal produces the bigram “really difficult,” thereby inflating the count for the “difficulty” sentiment type.

Example 2.8. Sample text showing original and prepared texts.

(original) No, really. Difficult as that may be...

(prepared) no really difficult as that may be

Lastly, and more generally, the absence of a testable baseline of established prior research weakens any claims in this research.

Recommendations

A number of significant aspects of this research are discussed in Section 2.1.4. Beyond those benefits, it is hoped that this research demonstrates the value of leveraging other disciplines. Many have been engaged in related linguistic research long before social media became a phenomenon. Social science, political science, linguistics, and psychology all include substantial bodies of theory relevant to opinion mining research. The concern of Liu (2012, p. 12) that, “Practical applications often demand more in-depth and fine-grained analysis" can best be met through a multi-disciplinary approach.

Proposition 2.1 introduces "sentiment-type" as an element to be included in the semantics of opinion. Similarly, Proposition 2.2 suggests, "sentiment-type-scale" is essential for opinion mining results to have meaning beyond positive or negative polarity. Further research is needed to explore these propositions.

Repeated studies on other large corpora are also needed to determine whether or not these findings are robust in different populations. More rigorous linguistic methods for the development of a sentiment taxonomy may produce better results. Additionally, Somasundaran (2010) developed an inventory of subjectivity types—or types of private states, which included some direct and indirect references to the sentiment types in U18. An inquiry into this relationship may be valuable.

An unexpected finding was that spam probabilities may correlate positively with sentiment trace densities. A manual review of documents which exceed statistical ranges for sentiment trace density showed high concentrations of SPAM. Future research may include using sentiment trace density as a way to identify SPAM content.

Lastly, as this research demonstrates that there is a strong commonality in sentiment expressions across languages in social media, future research may include development of a universal syntax for the structured expression of sentiment in text.

Acknowledgements

Special acknowledgment and thanks are due to Dr. Stephen Gilbert and Dr. Anthony Townsend who challenged me to answer the question: "Who says opinion is universal?"

BIBLIOGRAPHY

Abbasi, A., Chen, H., & Salem, A. (2008). Sentiment analysis in multiple languages: Feature selection for opinion classification in Web forums. *ACM Transactions on Information Systems (TOIS)*, *26*(3), 12.

Albig, W. (1957). Two decades of opinion study: 1936-1956. *Public Opinion Quarterly*, *21*(1), 14.

Baker, C., Fillmore, C., & Lowe, J. (1998). The berkeley framenet project. In *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics-Volume 1*, (pp. 86–90). Association for Computational Linguistics.

Bethard, S., Yu, H., Thornton, A., Hatzivassiloglou, V., & Jurafsky, D. (2004). Automatic extraction of opinion propositions and their holders. In *2004 AAAI Spring Symposium on Exploring Attitude and Affect in Text*, (pp. 2224).

Breck, E., Choi, Y., & Cardie, C. (2007). Identifying expressions of opinion in context. In *Twentieth International Joint Conference on Artificial Intelligence (IJCAI)*.

Burton, K., Java, A., & Soboroff, I. (2009). The icwsm 2009 spinn3r dataset. In *Proceedings of the Third Annual Conference on Weblogs and Social Media (ICWSM 2009)*.

Burton, K., Kasch, N., & Soboroff, I. (2011). The ICWSM 2011 Spinn3r Dataset. In *In Proceedings of the Fifth Annual Conference on Weblogs and Social Media (ICWSM 2011)*.

Chaffee, S. H. (1991). *Explication*, volume 1. Sage Publications, Incorporated.

Conrad, F. & Schober, M. (2007). *Envisioning the survey interview of the future*. Wiley-Interscience.

Davies, M. (2009). The 385+ million word corpus of contemporary american english (1990–2008+): Design, architecture, and linguistic insights. *International Journal of Corpus Linguistics*, *14*(2), 159–190.

De Melo, G. & Weikum, G. (2009). Towards a universal wordnet by learning from combined evidence. In *Proceeding of the 18th ACM conference on Information and knowledge management*, (pp. 513–522). ACM.

De Montaigne, M. (1580). *The complete essays of Montaigne*, volume 1.

Déjean, H., Gaussier, É., & Sadat, F. (2002). An approach based on multilingual thesauri and model combination for bilingual lexicon extraction. In *Proceedings of the 19th international conference on Computational linguistics-Volume 1*, (pp. 1–7). Association for Computational Linguistics.

Edmondson, D. (2005). Likert scales: A history. In *CHARM–the Conference on Historical Analysis and Research in Marketing*, (pp. 127–132).

Fu, G. & Wang, X. (2010). Chinese sentence-level sentiment classification based on fuzzy sets. In *Proceedings of the 23rd International Conference on Computational Linguistics: Posters*, (pp. 312–319). Association for Computational Linguistics.

Go, A., Huang, L., & Bhayani, R. (2009). Twitter sentiment analysis. *Entropy*, *17*.

Green, P. (1999). The 15-second rule for driver information systems. In *ITS America Ninth Annual Meeting Conference Proceedings*.

Hosp, B. & Vora, P. (2008). An information-theoretic model of voting systems. *Mathematical and Computer Modelling*, *48*(9-10), 1628–1645.

Hu, M. & Liu, B. (2004). Mining and summarizing customer reviews. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, (pp. 168–177). ACM.

Huang, T.-H. K. (2013). Social metaphor detection via topical analysis. In *Sixth International Joint Conference on Natural Language Processing*, (pp.14).

Indurkhya, N. & Damerau, F. J. (2012). *Handbook of natural language processing* (2 ed.). CRC Press.

Kamińska, D. & Pelikant, A. (2012). Recognition of human emotion from a speech signal based on plutchik's model. *International Journal of Electronics and Telecommunications*, *58*(2), 165–170.

Kim, S.-M. & Hovy, E. (2004). Determining the sentiment of opinions. In *Proceedings of the 20th international conference on Computational Linguistics*, (pp. 1367). Association for Computational Linguistics.

Kursten, J., Eibl, M., & der Nationen, S. (2008). Domain-Specific Cross Language Retrieval: Comparing and Merging Structured and Unstructured Indices.

Likert, R. (1932). A technique for the measurement of attitudes. *Archives of psychology*.

Liu, B. (2010). Sentiment analysis and subjectivity. *Handbook of natural language processing*, *2*, 627–666.

Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis Lectures on Human Language Technologies*, *5*(1), 1–167.

Macdonald, C., Ounis, I., & Soboroff, I. (2007). Overview of the TREC 2007 blog track. In *Proceedings of TREC 2007*.

Marcus, M., Santorini, B., & Marcinkiewicz, M. (1994). Building a large annotated corpus of English: The Penn Treebank. *Computational linguistics*, *19*(2), 313–330.

Miller, G. A., Beckwith, R., Fellbaum, C., Gross, D., & Miller, K. J. (1990). Introduction to wordnet: An on-line lexical database*. *International journal of lexicography*, *3*(4), 235–244.

Mishne, G. (2005). Experiments with mood classification in blog posts. In *Proceedings of ACM SIGIR 2005 Workshop on Stylistic Analysis of Text for Information Access*. Citeseer.

MIT, C. (2001). Voting: What is, what could be, report of the caltech mit voting technology project. Technical report.

Morgeson, F., Medsker, G., & Campion, M. (2006). A. Job and team design. *Handbook of Human Factors and Ergonomics (3rd Ed), Hoboken, NJ: John Wiley & Sons*, *428457*.

Novak, T. P., Hoffman, D. L., & Yung, Y.-F. (2000). Measuring the customer experience in online environments: A structural modeling approach. *Marketing Science*, *19*(1), 22–42.

Ortony, A. & Turner, T. J. (1990). What's basic about basic emotions? *Psychological review*, *97*(3), 315.

Osman, D., Yearwood, J., & Vamplew, P. (2007). Using corpus analysis to inform research into opinion detection in blogs, 65–75.

Pak, A. & Paroubek, P. (2010). Twitter as a corpus for sentiment analysis and opinion mining. In *LREC*.

Pang, B. & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, *2*(1-2), 1–135.

Pang, B., Lee, L., & Vaithyanathan, S. (2002). Thumbs up?: sentiment classification using machine learning techniques. In *Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10*, (pp. 79–86). Association for Computational Linguistics.

Piasecki, M., Marcinczuk, M., Musiał, A., Ramocki, R., & Maziarz, M. (2010). Wordnetloom: a graph-based visual wordnet development framework. In *Computer Science and Information Technology (IMCSIT), Proceedings of the 2010 International Multiconference on*, (pp. 469–476). IEEE.

Post, R. C. (1990). The constitutional concept of public discourse: outrageous opinion, democratic deliberation, and hustler magazine v. falwell. *Harvard Law Review*, 601–686.

Ptaszynski, M., Rzepka, R., Araki, K., & Momouchi, Y. (2012). Automatically annotating a five-billion-word corpus of japanese blogs for affect and sentiment analysis. In *Proceedings of the 3rd Workshop in Computational Approaches to Subjectivity and Sentiment Analysis*, (pp. 89–98). Association for Computational Linguistics.

Quirk, R., Crystal, D., & Education, P. (1985). *A comprehensive grammar of the English language*, volume 397. Cambridge Univ Press.

Reisenzein, R. (2009). Emotional Experience in the Computational Belief-Desire Theory of Emotion. *Emotion Review*, *1*(3), 214.

Rogers, E. M. (1976). New product adoption and diffusion. *Journal of consumer Research*, 290–301.

Shannon, C. (1948). A mathematical theory of communication. *Bell System Technical Journal*.

Somasundaran, S. (2010). *Discourse-level relations for Opinion Analysis*. PhD thesis, University of Pittsburgh.

Stenbro, M. (2010). A survey of modern electronic voting technologies.

Tang, L. & Liu, H. (2005). Bias analysis in text classification for highly skewed data. In *Fifth IEEE International Conference on Data Mining*, (pp.4).

Tang, L., Liu, H., Zhang, J., Agarwal, N., & Salerno, J. (2008). Topic taxonomy adaptation for group profiling. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, *1*(4), 1.

Tokuhisa, R., Inui, K., & Matsumoto, Y. (2008). Emotion classification using massive examples extracted from the web. In *Proceedings of the 22nd International Conference on Computational Linguistics-Volume 1*, (pp. 881–888). Association for Computational Linguistics.

Tsang, V. & Stevenson, S. (2010). A graph-theoretic framework for semantic distance. *Computational Linguistics*, *36*(1), 31–69.

Vagias, W. M. (2006). Likert-type scale response anchors. *Clemson International Institute for Tourism & Research Development, Department of Parks, Recreation and Tourism Management.*, 1p.

Van Steenburgh, E. (1987). Adverbial sensing. *Mind*, 376–380.

Wan, X. (2008). Using bilingual knowledge and ensemble techniques for unsupervised Chinese sentiment analysis. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, (pp. 553–561). Association for Computational Linguistics.

Wiebe, J. & Deng, L. (2014). An account of opinion implicatures. *CoRR*, *abs/1404.6491*.

Wiebe, J. & Mihalcea, R. (2006). Word sense and subjectivity. In *Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics*, (pp. 1065–1072). Association for Computational Linguistics.

Wilson, T., Wiebe, J., & Hoffmann, P. (2005). Recognizing contextual polarity in phrase-level sentiment analysis. In *Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing*, (pp. 354). Association for Computational Linguistics.

Winston, B. (1998). *Media technology and society: a history: from the telegraph to the Internet*. Psychology Press.

Xu, R., Wong, K., & Xia, Y. (2007a). Opinmine–Opinion Analysis System by CUHK for NTCIR-6 Pilot Task. In *Proceedings of the 6th NTCIR Workshop*.

Xu, R., Wong, K.-F., & Xia, Y. (2007b). Opinmine–opinion analysis system by cuhk for ntcir-6 pilot task. In *Proceedings of the 6th NTCIR Workshop*.

Zhai, Z., Liu, B., Wang, J., Xu, H., & Jia, P. (2012). Product feature grouping for opinion mining. *IEEE Intelligent Systems*, *27*(4), 0037–44.

Zheng, X., Lin, Z., Wang, X., Lin, K.-J., & Song, M. (2014). Incorporating appraisal expression patterns into topic modeling for aspect and sentiment word identification. *Knowledge-Based Systems*, *61*, 29–47.

CHAPTER 3. SPEQ-ING THE TRUTH: THE STATES, PROCESSES, EFFECTS, AND QUALITY MODEL FOR OPINION MINING AND SENTIMENT ANALYSIS

In preparation for submission to the International Conference on Social Computing 2015.

Erin Mikel Phillips[1]

Abstract. Opinions are private states. Votes are formal expressions of opinion. Opinion mining research aims to make informal expressions of opinion in text, formal, by giving them structure through algorithmic inference. However, opinion mining research is not informed by a comprehensive theoretical model of opinion which integrates the concepts of private states, opinion, and voting. This paper derives the States, Processes, Effects, and Quality (SPEQ) Model for Opinion Mining and Sentiment Analysis using a qualitative review of social psychology, opinion mining, and voting systems research. SPEQ includes an end-to-end model of opinion from a holder's belief and desire through reporting by an aggregator. SPEQ defines seven possible states for opinion, six processes which govern a transition between those states, and five quality and integrity measures which instrument those processes. Also, the term "voot" is introduced as a verb to represent a formal expression of opinion in an informal context, such as social media. SPEQ has the potential to enable a more consistent and fine-grained analysis of opinion mining and sentiment analysis research.

[1]Primary researcher and author.

3.1 Introduction

With the global adoption of social media platforms, individuals and organizations can express themselves and consume the expressions of others in ways not conceivable as recent as ten years ago. In 2004, for example, Facebook was called, "thefacebook.com", and Twitter was two years from becoming an idea. Today, these two brands alone daily serve billions of personal, professional, educational, entertainment, legal, governmental, and other sundry moments and messages to the social media eco-system. Much of the content generated and consumed in social media contains expressions of opinion. Macdonald et al. (2007) showed that approximately 50% of the textual content in social media is opinion-laden, and this figure was confirmed by Phillips (2011). The vast quantities of opinion expressed in social media have fueled a corresponding surge in research in opinion-related research. The emergence of opinion mining and sentiment analysis as a research discipline has largely been a result of that surge. The relationship between the proliferation of social media platforms and the growth of opinion mining and sentiment analysis research has been symbiotic. The pace of opinion expression in social media and the race to reliably consume those expressions have both been frenetic.

A brief look at the statistical character of these two phenomena is included in this introduction. No citations were available which addressed this relationship. Therefore, a quantitative digression is warranted in what otherwise is a qualitative study. The purpose of this quantitative digression is to demonstrate the numerical character of the two phenomena and the closeness of the relationship between them.

Mathematically, the rate of publication of opinion mining and sentiment analysis research is on an exponential growth curve. The trend for opinion mining related publications is approximately $23 * e^{0.44(YEAR-1999)}$ ($R^2 = 0.97$.)

Operationally, it can be reasonably postulated that the rate of adoption and diffusion of social media platforms and the rate of publication on opinion-related research are positively correlated. This proposition can be stated formally in the form of Null Hypothesis 3.1.

Null Hypothesis 3.1. *For the years 2000 to 2014, the number of scholarly articles published on opinion mining and sentiment analysis in a given year is independent of the number of Facebook users for the same year.*

The relationship between social media adoption and published opinion mining and sentiment analysis scholarship is shown graphically in Figure 3.1. Statistically, there is a strong correlation (p<.001 using Prais-Winsten, $R^2 = 0.7$.) We can confidently reject Null Hypothesis 3.1. The number of articles published on opinion mining and sentiment analysis and the number of Facebook users are related.

The article counts were produced using Google's Scholar search engine. The search term "airplane" was included as a control. This inclusion was done to confirm that no systemic artifact was present. The Google Scholar service itself was only introduced in 2004 and has undergone many changes to date. Its coverage is competitive with other scholarly search engines (Kulkarni et al., 2009).

The line representing opinion mining research on the graph shown in Figure 3.1 begs the question, "what is so interesting and challenging about opinion mining and sentiment analysis that motivates such a prolific rate of scholarly inquiry?" No single answer will do the question full justice. However, Liu (2012), in his comprehensive survey outlines the vision and promise of opinion mining and sentiment analysis research:

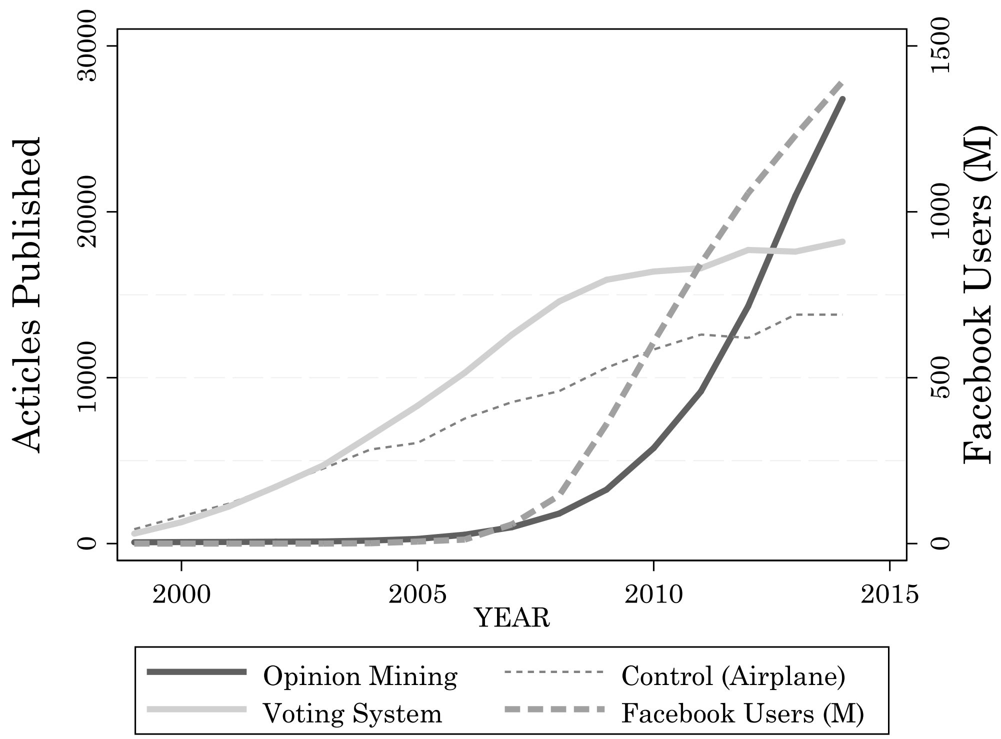


Figure 3.1: Opinion mining and sentiment analysis, voting systems, and airplane-related publication counts by year compared to Facebook user counts for the same year

> "it may no longer be necessary to conduct surveys, opinion polls, and focus groups in order to gather public opinions because there is an abundance of such information publicly available ... to almost every possible domain, from consumer products, services, healthcare, and financial services to social events and political elections." (p. 8-9)

The scope of the previous quote is both substantial and significant. Moreover, the quote from Liu (2012) sets the bar high.

However, it is important to keep in mind that it is the informal and unstructured expressions of social media users that is the "information publicly available." Not, opinion. "Text" is what is publicly available—and this distinction is critical. As Tang, Tan & Cheng (2009, p. 10760) writes, "The goal ... is to identify direct and indirect sources of opinions, emotions, sentiments, and other private states that are expressed in text."

According to Liu (2012) the functional goal of opinion mining research is to establish opinion mining as a better method of counting opinions. Such a method would supersede other more conventional opinion counting systems (i.e., social science methods.)

According to Tang et al. (2009), the technical goal of opinion mining and sentiment analysis research is to explore mechanisms and methods which can extract private states (including opinions) from textual discourse. Achieving this would achieve the functional goal of Liu (2012).

Both Tang et al. (2009) and Liu (2012) describe opinion mining and sentiment analysis as a way to formalize (give structure and specific meaning) to opinions written in free text. Without resorting to extrapolation the following proposition naturally follows:

Proposition 3.1. *A primary goal of opinion mining and sentiment analysis research is to identify ways to formalize and make structured what are otherwise informal and unstructured expressions of opinion in social media.*

The formalizing and giving structure to opinion expressions in social media, as indicated in Proposition 3.1, is what is so intriguing and challenging.

Evidence of the empirical emphasis is abundant. The vast majority of the research 73 papers published in what Pang & Lee (2008, p. 7) called "the sentiment analysis and opinion mining ... land rush" are empirical studies. These develop then evaluate quantitative methods which move closer to the goal of Liu (2012).

As for theoretical expositions, there have been a few attempts. However, these take the form of compilations of recent empirical works (Pang & Lee, 2008; Liu, 2012; Othman, Hassan, Moawad & El-Korany, 2014).

Another aspect of the bias toward empirical work is that it is not clear that researchers see gaps in understanding which call for new cross-cutting theories. As Liu (2012, p. 14) describes the state of opinion mining and sentiment analysis research, "Due to the maturity of the field... the problem [opinion mining and sentiment analysis] is now better defined and different research directions are unified around the definition."

The situation facing opinion mining and sentiment analysis research is strikingly reminiscent of that observed by Albig (1957) in his retrospective on 20 years of public opinion research since the advent of polling:

> "During the past twenty years, several thousand articles dealing with public opinion, the mass media, and communications have been published. Polling, attitude measurement, and market research have become an industry expending probably one hundred million dollars a year. ... In spite of this enormous production, I am not encouraged when I review what I have learned of meaningful, theoretical significance about communications and what I have learned about the theory of public opinion." (p. 14-15)

Albig goes on to point to Lazarsfeld & Katz (1955) as an example of the kind of "daring generalization" which would advance understanding in public opinion research. Interestingly, Lazarsfeld's work was a decade or more in the making when referenced by Albig. However, both social media and the scholarly disciplines trying to make sense of social media have been rapidly growing and changing. Comprehensive theoretical models which describe the relationships among the core concepts have been difficult if not impossible to develop. Notwithstanding the obstacles, it is Albig's challenge to researchers in Public Opinion and Communications which frames this inquiry.

Therefore, the object of this paper is to make some "daring generalizations." The approach is to take a holistic look at opinion mining and sentiment analysis research in the context of other related disciplines. These other related disciplines include social psychology and voting systems research. The goal is to develop a richer theoretical framework for opinion mining and sentiment analysis research around the more general concepts of opinion and social capital.

3.1.1 Problem

The noun phrase "voting system" was included Figure 3.1 because the words "opinion" and "vote" have a strong semantic relationship. At this point, definitions for "vote" and "opinion" would be helpful.

Each of these terms will be defined more rigorously in Section 3.2.5, but for purposes of framing the problem, the following definitions are sufficient. The definitions of "vote" and "opinion" shown in Definitions 3.1 and 3.2 come from Merriam-Webster (2004).

Definition 3.1. Vote: a usually formal expression of opinion.

Definition 3.2. Opinion: a view, judgment, or appraisal formed in the mind [a private state] about a particular matter.

Of course, Definitions 3.1 and 3.2 make the semantic relationship between "opinion" and "vote" obvious. The latter is a "usually formal" expression of the former.

The motivation for including the term "voting system" in Figure 3.1 is derived from this semantic relationship. The definition of "voting system" shown in Definition 3.3 extends the parallelism between "opinion" and "vote" to "opinion mining" and "voting systems." This definition comes from Hosp & Vora (2008).

Definition 3.3. Voting system: a vote counting system.

This paper intends to demonstrate that the semantic relationship between "opinion" and "vote" has significant potential for clarifying and extending the rhetoric around opinion mining and sentiment analysis.

As demonstrated above, Liu (2012) and Tang et al. (2009) see opinion mining and sentiment analysis as a way to create a voting system from textual discourse via the automated extraction of opinions. However, opinions are private states (Post, 1990).

Polls, surveys, and focus groups are no longer as fascinating after the election because all the votes are cast. By analogy, the same is true for polls, surveys, and focus groups if the Liu (2012) aspirations are achieved. Sufficient information, in the form of formalized and structured representations of informal and unstructured expressions of opinion, would be publicly, freely, and abundantly available.

Given the above definitions and rationale, the following two claims are available, and their implications for the problem statement which defines the boundaries of this paper. The first claim is immediately below. The second claim will follow.

Claim 3.1. Opinions are private and invisible.

Post (1990) provides a legal definition of opinion. This definition aligns itself with private states—visible only to the individual—as being the essence of subjectivity and the defining characteristic of opinion.

> "[opinions] are not objectively verifiable or subject to empirical proof... for constitutional purposes the truth of certain kinds of statements — *opinions* — can only be determined by the free play of speech and counterspeech characteristic of the marketplace of ideas." (p. 656)

In the context of this research, the "free play of speech and counterspeech" in the above definition is social media. What is measured in opinion mining and sentiment analysis is the encoded expression of private state information—not the opinion itself.

This paper asserts that the focal concept of opinion mining and sentiment analysis research is a private state. However, a private state is not directly measurable. Interestingly, this relationship is seldom if ever discussed in opinion mining and sentiment analysis literature. A full-text query and subsequent manual review of usage of the over 200 publications showed no theoretical or operational model which accounted for the chain of custody of an opinion from private state to decoded representation. It is this lack of accounting for the essence of opinion which fuels the definition of the first major problem addressed by this paper.

Problem 3.1. Though Claim 3.1 asserts the private state nature of opinion, opinion mining and sentiment analysis research is not conducted within a theoretical framework which comprehends this fact.

The second claim formalizes the relationship between terms "opinion" and "vote."

Claim 3.2. Opinions expressed in a structured manner are "votes."

The language of Claim 3.2 differs slightly from the Definition 3.2 in that "structured" replaces "formally." The clarification seems small but warranted. The word "formally" emphasizes an accord with conventions rather than the operative criteria for a vote—an accessible semantic structure, which creates the potential to be counted. At this point, Null Hypothesis 3.1 can be adapted to look at the relationship between opinion mining and voting systems scholarship.

Null Hypothesis 3.2. *The number of scholarly articles published on voting systems in a given year is independent of the number of articles published on opinion mining and sentiment analysis for the same year, for the years 2000 to 2014.*

The rejection of Null Hypothesis 3.1 seemed self-evident given the context and proximity of the concepts to social media. Whether or not Null Hypothesis 3.2 can be rejected seems less clear. Context and semantic proximity are as helpful. Looking at Figure 3.1, however, the answer is equally apparent though with a different outcome. The two concepts do not appear to be related.

It intuitively seems that there should be some relationship. The definitions for “vote”, “opinion”, and “voting system” in Definitions 3.1 to 3.3 suggest a relationship. The finding, however, is that there is no detectable relationship. The opinion mining and sentiment analysis rate of scholarly publication and that of the voting systems are not correlated (p=1 using Prais-Winsten, $R^2 = neglig.$) Null Hypothesis 3.2 is not rejected.

As a qualitative check of this finding, a full-text query and subsequent manual review of relevant literature showed no meaningful matches for the words “vote” or “voting” as a focal concept. The terms “vote” and “voting” are sometimes found, but mainly in the context of the methodology employed for establishing “ground truth” as a baseline for accuracy of a sentiment classifier. Liu (2012) references the term “vote” only in this manner. Liu (2012, p. 138) states, “the helpfulness *votes* as the ground truth may not be appropriate because of [3] biases”

Voting systems scholarship is dedicated to defining methods for measuring and improving the quality of voting systems—systems for counting votes according to Definition 3.3. Opinion mining and sentiment analysis scholarship is dedicated to defining methods for measuring and improving the quality of less formal voting systems through textual analysis according to Claim 3.2. It is curious then, that the terms "vote", "voting", and "voting system" are not found in any meaningful context in opinion mining and sentiment analysis scholarship.

It is this lack of relatedness between the two fields which forms the definition of the second problem this paper seeks to address. A few preconditions are listed below, followed by a formal problem statement.

- Claim 3.2 substantially asserts that there is a strong semantic relationship between the words opinion and vote.
- Liu (2012) and Tang et al. (2009) explain that the goal of opinion mining and sentiment analysis research is to transform unstructured opinions into "votes" which can be counted.
- Definition 3.3 tells us that the purpose of voting systems scholarship is to better understand and implement systems of counting votes.

Problem 3.2. *Opinion mining and sentiment analysis research is not conducted within a theoretical framework which comprehends the strong semantic relationship between "opinion" and "vote"—and the strong operational similarities between "opinion mining systems" research and "voting systems" research.*

3.1.2 Purpose

Problems 3.1 and 3.2 encapsulate the finding discussed above that the theoretical and practical relationship between "opinion" and "vote" is essentially unexplored in opinion mining and sentiment analysis literature and methodology. This apparent gap in the literature and research is the impetus for this research.

The general purpose of this paper is to expand the narrative around opinion mining and sentiment analysis research: to the make more tangible the connection between the field of opinion mining and sentiment analysis research and both the psychology of opinion as "private states" and voting systems scholarship.

Aside from the empirical analysis already presented, this is a qualitative paper which seeks to generate by induction an expanded theoretical model for opinion mining and sentiment analysis. This objective model should support Proposition 3.1, account for the semantics in Definitions 3.1 to 3.3, affirm Claims 3.1 and 3.2, and ultimately be meaningfully responsive to Problems 3.1 and 3.2.

Research Questions

The need for a comprehensive model of opinion was in Section 3.1.1. Throughout the development of a new or expanded model of opinion mining and sentiment analysis research, the following research questions will serve as guides. The first research question evaluates the impact of Problem 3.1.

Research Question 3.1. *What impacts, if any, are there on opinion mining and sentiment analysis research if Claim 3.1, "opinions are private", is accepted and operationalized?*

The second research question is similar, but focuses on the relationship between opinion mining and voting systems research.

Research Question 3.2. *What impacts, if any, are there on opinion mining and sentiment analysis if Claim 3.2, that "structured representations of opinion are votes", is accepted and operationalized?*

The similarity in structure of Research Questions 3.1 and 3.2 portends a synergistic outcome: a single theoretical model which connects "private states" to opinion mining and sentiment analysis research, and then extends the model to incorporate the formalisms of voting systems scholarship.

3.1.3 Approach

A multi-disciplinary approach is needed to meet the purposes discussed in Section 3.1.2. Therefore, this work is a hybrid qualitative and quantitative analysis of social psychology, opinion mining and sentiment analysis, and voting systems scholarship.

The methodology used in the course of this research is an innovative adaptation of Qualitative Meta-Synthesis (QMS). QMS is a technique often used in medical research to develop theoretical models from collections of qualitative case studies (Walsh & Downe, 2005).

The goals of QMS include theory development, which aligns closely with those of this research. However, the subject matter is different. In this case, the case studies represent published scholarship in social psychology, opinion mining and sentiment analysis, voting systems, and other related disciplines.

The QMS process is a multi-step process which moves through a number of stages. QMS starts with concept inventories. Then, there is concept alignment—identifying flows and relationships. Lastly, there is a consolidation of terms. The literature for this particular QMS analysis was selected from a population of related research papers, books, and journals. The selection was done using customized linguistic analysis tools developed specifically for the purpose (CiteScan, see Appendix A.)

3.1.4 Significance

This paper is an attempt to advance the field of opinion mining and sentiment analysis through a qualitative review of a wide span of intellectual inquiry. It has the potential to be widely influential, because it is the first of its kind.

The projection of value assumes that the basic premises about weaknesses in existing theoretical and operational models are proved correct. Moreover, it assumes that the new model resulting from the QMS process in this research proves to be robust.

The model developed in the course of investigating Research Questions 3.1 and 3.2 is the States, Processes, Effects, and Quality (SPEQ) Model for Opinion Mining and Sentiment Analysis Research.

Because SPEQ is an end-to-end theoretical model, it has the potential to provide a framework which coalesces a difficult lexical and conceptual space. The adoption process for SPEQ would likely include the introduction of new words, such as "voot", in response to provable gaps in the lexical landscape around opinion mining and sentiment analysis.

3.2 Background

This section contains a review of definitional work relevant to Research Questions 3.1 and 3.2.

Within the field of opinion mining and sentiment analysis, published research is dominated by empirical studies. These evaluate algorithmic approaches to extracting meaning from free text. No generally accepted theoretical framework binds these many empirical studies. Therefore, it was a challenge to conduct a literature review for this research. The number of interdisciplinary and theoretical works available for review is very small. Of course, this gap is one of the key motivations for this qualitative literature analysis.

Two important elements related to opinion are not represented in opinion mining and sentiment analysis scholarship. These are "private states" and the "voting-systems-like" character of opinion mining and sentiment analysis. These two concepts and the concept of opinion itself are foundational to this investigation. The

QMS theory development methodology adapted for this paper begins with these concepts.

3.2.1 Diligence

The methodology of this literature review is adapted from Qualitative Meta-Synthesis (QMS). QMS is a technique often used in medical research to develop theoretical models from collections of qualitative case studies (Walsh & Downe, 2005).

A hybrid instantiation of the QMS processes was used because the scope of this research spans multiple disparate disciplines. Corpora-size libraries of related literature also called for natural language process (NLP) methods.

The goals of QMS include theory development, which aligns closely with those of this research—albeit in a different, and non-medical literature domain. In this case, the case studies represent published scholarship in social psychology, opinion mining and sentiment analysis, voting systems, and other related disciplines. A substantial effort was made to extend CiteScan to include quantitative methods adapted from corpus linguistics to identify sources of influence in relevant papers—which would inform the QMS process. The adapted QMS process followed for this research includes the following steps:

1. Define the scope of the inquiry.
2. Determine the population of source documents relevant to the research.
3. Identify nominal concepts which may be relevant within the selected research papers.
4. Collapse and consolidate definitions, using the individual or cultural constructs within the selected research papers.
5. Identify the relationships among the concepts across the portfolio of selected research papers.

6. Translate the concepts and relationships across the portfolio, to develop a common representation.

7. Synthesize the resulting translations to, as Walsh & Downe (2005, p. 209) explains, "elucidate more refined meanings, exploratory theories, or new concepts."

Steps 1-3 constitute the balance of this section; steps 4-6 constitute Section 3.3, the section on Analysis. Lastly, step 7, Synthesis, is Section 3.4, where the final theory development work is done.

A number of customizations to the QMS process have been implemented for this research. It may be inexperience which motivated these extensions or alterations, as the primary investigator for this paper had not previously used QMS prior to undertaking to use it as a methodology for this qualitative inquiry. QMS customizations are called out in the context of the relevant step. So, while the findings produced are expected to be useful—the specific techniques used in the application of QMS may be of limited value. Walsh & Downe (2005) provides a useful overview of QMS in the context of medical scholarship.

Quantitative Methods. The diversity of subject matter relevant to this research prompted the development of some tools to enable bibliographical analysis of the literature. Where these tools were somehow definitive in their application, they are noted; in all other ways, the tools were either not used or used in an ancillary fashion. Details of the CiteScan tool's design are provided in Appendix A.

3.2.2 Disclosures

No human subjects were used in the execution of this study. This research was otherwise conducted in accordance with the guidelines published by Iowa State University Institutional Review Board regarding the protection of human participants in the Investigator Handbook.

3.2.3 Scope

In QMS, the outcome of the scoping step is a few words or phrases which include the elements necessary to the inquiry—optionally with some relational modifiers. Little specific guidance is provided within QMS regarding the scoping of an inquiry. The QMS scoping criterion used in this paper is open-ended. Walsh & Downe (2005, p. 206) prescribes, "[the scope] must allow for ... [refutation] which come to oppositional conclusions from the main body of the work in a particular area."

Research Questions 3.1 and 3.2 set the direction for this inquiry and are a good place to start. They meet the "refutation" criteria. The justification for these research questions in Section 3.1 is derived from an apparent divergence between current theoretical models and those needed to adequately described and analyze the relevant phenomena. Therefore, the scope of the QMS process followed for this paper will be set to align with the Research Questions 3.1 and 3.2.

Adequately responding to the research questions is the outcome of the QMS process itself. Therefore, simpler scoping statements are used to indicate the core elements fundamental to the inquiry. Rather than using two separate scoping statements—which may unnecessarily segregate the concerns of Research Questions 3.1 and 3.2, this investigator chose to develop a single integrated scoping statement.

For Research Question 3.1, a scoping statement could be defined as: "private states, as opinions." For Research Question 3.2, a valid scoping statement could be: "opinion expressions, decoded as votes, counted by a voting system." The single conjoined statement used in this inquiry is shown below in the form of Claim 3.3:

Claim 3.3. *Private states inform opinion expressions which may be counted by a voting system.*

QMS customizations in this step include the use of a claim as the scoping statement. It was felt that the affirmative nature of a claim would ensure sufficient semantic depth in the scoping statement. Also, the use of the term “portfolio” does not appear in the QMS literature which was reviewed in preparation for this study. The term ”portfolio” is used to reference the population of papers selected for the QMS process.

3.2.4 Population

The population step in QMS involves the search for and selection and appraisal of research papers to be included in the analysis. The best practice in QMS for this step, according to Walsh & Downe (2005, p. 206) is to, “undertake a robust search on the topic area as one would do in the early stages of undertaking a systematic review.”

The systematic review undertaken of scholarship relevant to Claim 3.3 included important surveys of social psychology, opinion mining and sentiment analysis, and voting systems scholarship. Moreover, some deep-dive or canonical representation papers were identified to improve the coverage of important topics.

The following method was used to identify the articles to be considered. The purpose of the document selection is not to be comprehensive for any particular domain—but to be representative. The expectation is that across a portfolio of publications, important thematic elements will be present—and the search algorithms will help promote documents with higher relevance.

Method. The QMS process for selecting a population of sufficiently relevant documents is not well defined—but is characterized as an iterative process. Walsh & Downe (2005, p. 207) explained that the population of relevant documents includes, "iteration around the scope of the review . . . [through and] until the final stages of the synthesis."

For purposes of this research, it was decided to use the inventory of documents already identified and indexed to date. The alternative was to generate a separate index, or rely strictly on a search engine. However, these were rejected over quality and breadth concerns.

The initial inventory of documents considered includes 912 papers, books, and articles, in a single repository. These 912 documents constitute the starting point, the "raw portfolio" of documents for use in the QMS process.

First pass, indexability. The first pass over the raw portfolio is intended to ensure a conforming set; that is, that all documents are equally searchable. Of the 912 documents, 893 could be converted to UTF-8 text for indexing without modification. The text of the remaining 893 documents were loaded into a simplified version of CiteScan (no bibliographical indexing), for analysis.

Second pass, relevance. The second pass over the raw portfolio is intended to define a subset of these documents, the most relevant ones, while avoiding specialized types of selection which could lead to narrow interpretations.

The queries used within CiteScan were derived from the text of Claim 3.3. Large documents have a built-in bias for general measures of relevancy because there is a larger pool of words available. Two relevance algorithms were applied to prevent the domination of the relevance scores by large documents. The first was unweighted—where the number of words in the document didn't matter. The second discounted relevance scores logarithmically for document size.

CiteScan returned 71 out of 893 documents with relevance scores greater than 1.0 using a search criteria derived from the text of Claim 3.3. The cutoff value of 1.0

was selected somewhat arbitrarily—to focus the QMS process on the most relevant set of documents and to keep the size of the set of documents manageable.

This initial set of 71 documents was selected based upon a general query. The following algorithm was used by CiteScan to reduce the list of 71 to a smaller set of the most impactful documents relative to Claim 3.3. The algorithm to identify the documents to include is described below in simplified terms. The symbol Q represents the query string from Claim 3.3:

1. Using Q, find documents with the highest relevance scores.
2. Using the meaningful n-grams (n=1,2) from Q, find all documents with non-zero relevance scores.
3. Using the n-grams from #2, find all documents with non-zero relevance scores, adjusting those scores by document word count.
4. Trim each of the resulting lists from steps 1-3 to 25, to ensure a balance between completely general relevance (step 1), large-document relevance for specific n-grams (step 2), and concentrated document relevance for specific n-grams (step 3).
5. For each of the lists in step 4, calculate a normalized index score for each document as follows: $(26-rank_{relevant})*rele\hat{v}ance$, where $rele\hat{v}ance$ is a normalized value between 0 and 1.
6. Combine all documents from lists in step 5, and calculate a new index score using the sum of all list-specific scores for each document.
7. Sort the remaining list of documents in descending order by cumulative index score.
8. Return the set of documents which constitute 95% of the distribution of index scores.

When CiteScan had completed applying this algorithm, the resulting portfolio included 25 documents.

Third pass, other factors. Because the purpose of this research is to generalize, it was felt that a purely mechanized selection of documents could have some bias not visible to the investigator which might constrain the inquiry. Therefore, two additional document selections were included.

First, it was decided to select five documents at random from the 75 documents which survived the first pass, but were pruned in the second pass. The purpose of "sprinkling in" a small random set of papers which were otherwise excluded was to reach beyond any "algorithmic orthodoxy." This approach forced the inclusion and review of research approaches and outcomes not otherwise anticipated as being highly relevant.

Secondly, it was decided to allow the investigator to add a few documents from the entire raw portfolio of 893 documents, which, by the investigator's judgement, should have been included but weren't. This decision was done a-priori, with the full knowledge that there would be many documents which met this criteria when CiteScan was done with the second pass. Interestingly, both documents selected by the investigator were in the set of 75 documents which survived the first pass, but not the second.

The Final QMS Document Portfolio. The resulting final portfolio, derived from the raw portfolio of 912 documents, included:

- 25 documents which had the highest relevance to the text of Claim 3.3.
- 5 documents selected at random from the 75 excluded from the second pass.
- 2 documents selected by the investigator from the raw portfolio of 912 documents which were not otherwise included in the final portfolio.

The 32 documents which make up the final QMS portfolio of documents for the rest of the process is shown in Table 3.1. For a completeness and bias check, the

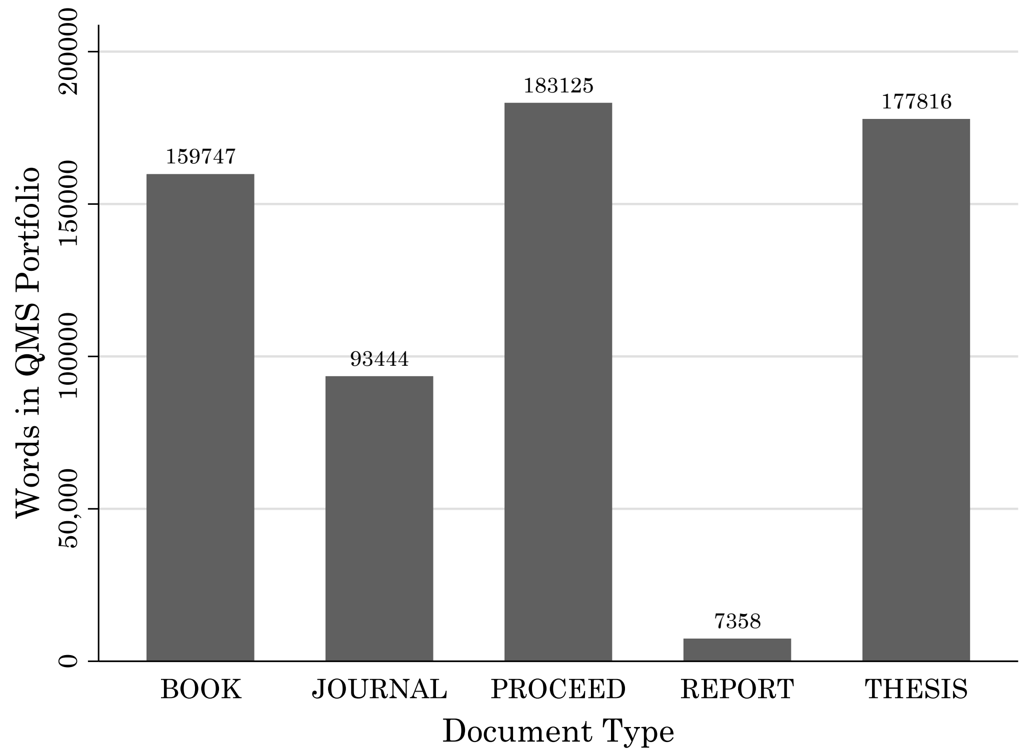


Figure 3.2: QMS portfolio word counts by document type

composition of these 32 documents is shown by type in Figure 3.2, and by year in Figure 3.3.

3.2.5 Concepts

The purpose of this section is to extract a set of nominal concepts from the portfolio—with as little interpretation and extension as possible. In simple terms, answer the question, "What is in this collection of documents?"

As Walsh & Downe (2005, p. 208) describes, the purpose is to collect, "the original author's understanding of key metaphors, phrases, ideas, concepts, and relations in each study. . . and is completed with the creation of a grid of key concepts."

This paper adopts a similar strategy, leveraging CiteScan outputs as a way to chart a reliable course through the 1600+ pages of scholarship in the 32 documents included in the portfolio. Having employed a holistic selection process with CiteS-

Table 3.1: QMS portfolio of 32 documents

DOCUMENT		TYPE	REASON	WORDS	PRIVATE STATE	OPINION	VOTING SYSTEM
Akkaya (2013)	*Subjectivity Word Sense Disambiguation: A Tool For Sense-Aware Subjectivity Analysis*	T	RANDOM	38,677	8	38	-
Alvarez & Nagler (2000)	*Likely Consequences Of Internet Voting For Political Representation, The*	J	RELEVANCE	12,692	-	3	1
Alvarez et al. (2008)	*Internet Voting In Estonia*	R	RELEVANCE	7,358	-	-	9
Appel et al. (2009)	*The New Jersey Voting-Machine Lawsuit And The Avc Advantage Dre Voting Machine*	P	RELEVANCE	12,955	-	2	10
Bethard et al. (2004)	*Automatic Extraction Of Opinion Propositions And Their Holders*	P	RELEVANCE	6,115	-	232	-
Ding et al. (2008)	*A Holistic Lexicon-Based Approach To Opinion Mining*	P	RELEVANCE	8,430	-	233	-
Feldman & Benaloh (2009)	*On Subliminal Channels In Encrypt-On-Cast Voting Systems*	P	RELEVANCE	7,458	-	-	24
Hall (2006)	*Transparency And Access To Source Code In Electronic Voting*	P	RELEVANCE	12,176	-	-	149
Hosp & Vora (2008)	*An Information-Theoretic Model Of Voting Systems*	J	RELEVANCE	13,549	-	1	155
Kim & Hovy (2006)	*Extracting Opinions, Opinion Holders, And Topics Expressed In Online News Media Text*	P	RELEVANCE	40,744	1	403	-
Kisilevich et al. (2010)	*Beautiful Picture Of An Ugly Place: Exploring Photo Collections Using Opinion And Sentiment Analysis Of User Comments*	P	RANDOM	8,201	-	140	-
Liu (2012)	*Sentiment Analysis And Opinion Mining*	B	RELEVANCE	67,802	-	1025	-
Loncke & Dumortier (2004)	*Online Voting: A Legal Perspective*	J	RELEVANCE	11,554	-	9	57
Luskin et al. (1999)	*Deliberative Polling, Public Opinion, And Democracy: The Case Of The National Issues Convention*	P	MANUAL	11,692	-	6	-
NASED (2002)	*Voting System Standards, Volume 1*	P	RELEVANCE	45,573	-	-	365
Pitt et al. (2005)	*Formalization Of A Voting Protocol For Virtual Organizations*	P	RELEVANCE	7,045	-	1	3
Popoveniuc et al. (2010)	*Performance Requirements For End-To-End Verifiable Elections*	P	RELEVANCE	13,997	-	-	90
Prevost & Schaffner (2008)	*Digital Divide Or Just Another Absentee Ballot? Evaluating Internet Voting In The 2004 Michigan Democratic Primary*	J	RELEVANCE	8,077	-	-	-
Rivest & Smith (2007)	*Three Voting Protocols: Threeballot, Vav, And Twin*	P	RELEVANCE	9,880	-	-	37
Somasundaran (2010)	*Discourse-Level Relations For Opinion Analysis*	T	RELEVANCE	71,406	16	979	-
Stark (2010)	*Super-Simple Simultaneous Single-Ballot Risk-Limiting Audits*	P	RELEVANCE	7,894	-	-	1
Stenbro (2010)	*A Survey Of Modern Electronic Voting Technologies*	B	RELEVANCE	46,372	-	4	252
Svensson & Leenes (2003)	*E-Voting In Europe: Divergent Democratic Practice*	J	RELEVANCE	7,513	-	1	3
Tang et al. (2009)	*A Survey On Sentiment Detection Of Reviews*	J	RANDOM	15,265	1	166	-
Teague et al. (2008)	*Coercion-Resistant Tallying For Stv Voting*	P	RELEVANCE	10,306	-	-	7
Wiebe & Deng (2014)	*An Account Of Opinion Implicatures*	J	RELEVANCE	19,515	78	82	-
Wiebe et al. (2005)	*Annotating Expressions Of Opinions And Emotions In Language*	J	MANUAL	12,343	58	62	-
Wilson (2008)	*Fine-Grained Subjectivity And Sentiment Analysis: Recognizing The Intensity, Polarity, And Attitudes Of Private States*	T	RELEVANCE	67,733	248	99	-
Xu et al. (2007)	*Opinmine--Opinion Analysis System By Cuhk For Ntcir-6 Pilot Task*	P	RELEVANCE	6,495	-	204	-
Zhang & Liu (2011)	*Identifying Noun Product Features That Imply Opinions*	P	RELEVANCE	3,796	-	141	-
Zhang & Ye (2008)	*A Generation Model To Unify Topic Relevance And Lexicon-Based Sentiment For Opinion Retrieval*	P	RANDOM	5,949	-	119	-
Zhang et al. (2007)	*Opinion Retrieval From Blogs*	P	RANDOM	9,992	-	201	-

T=THESIS; J=JOURNAL; P=PROCEEDINGS; B=BOOK; R=REPORT

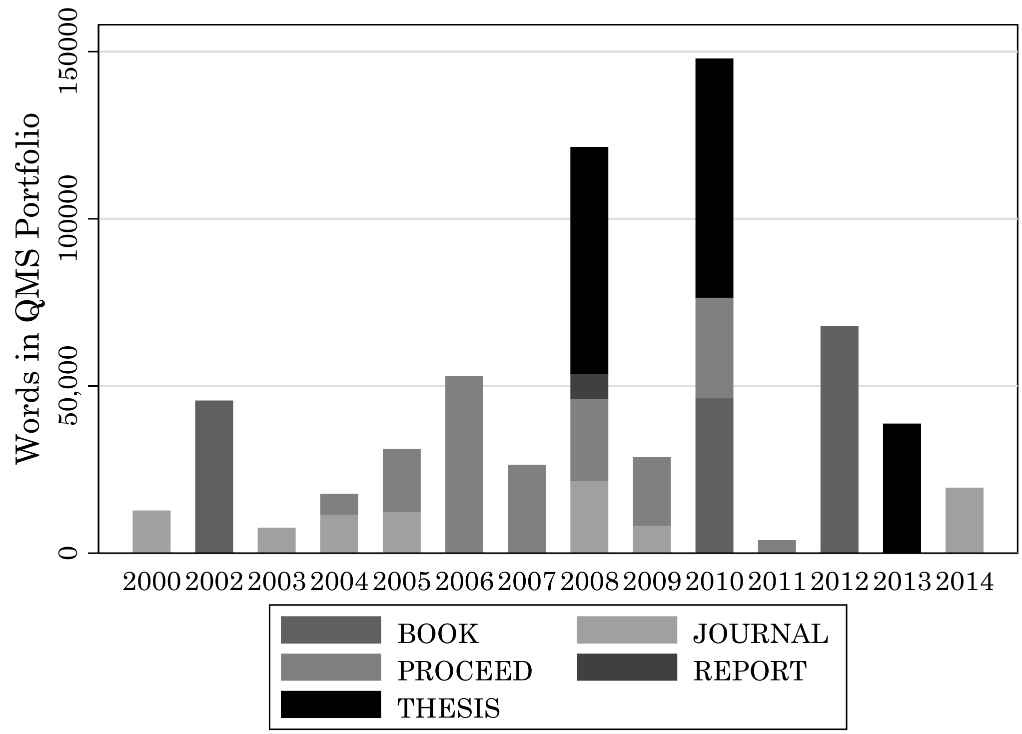


Figure 3.3: QMS portfolio word counts by year and document type

can in the previous section, Section 3.2.4, the next step is to take a look at the documents in the portfolio. Cluster analysis was used to establish how the documents are related. This process guides the extraction of nominal concepts from the documents in the portfolio. With such a large quantity of material to digest, consuming related materials at the same time will help see cross-cutting concerns more efficiently.

CiteScan was used to extract word frequencies for "private state", "opinion", and "voting system." These counts are included at the end of Table 3.1. Simple clustering was done to try to group the papers in the portfolio based upon the word frequencies. The cluster relationships are shown in a dendritic graph in Figure 3.4.

Also, a word frequency map was also created for each document in the final portfolio. Figure 3.5 shows an example of the word frequency map for Stenbro (2010). A complete set of word maps is included in Appendix B because each offers

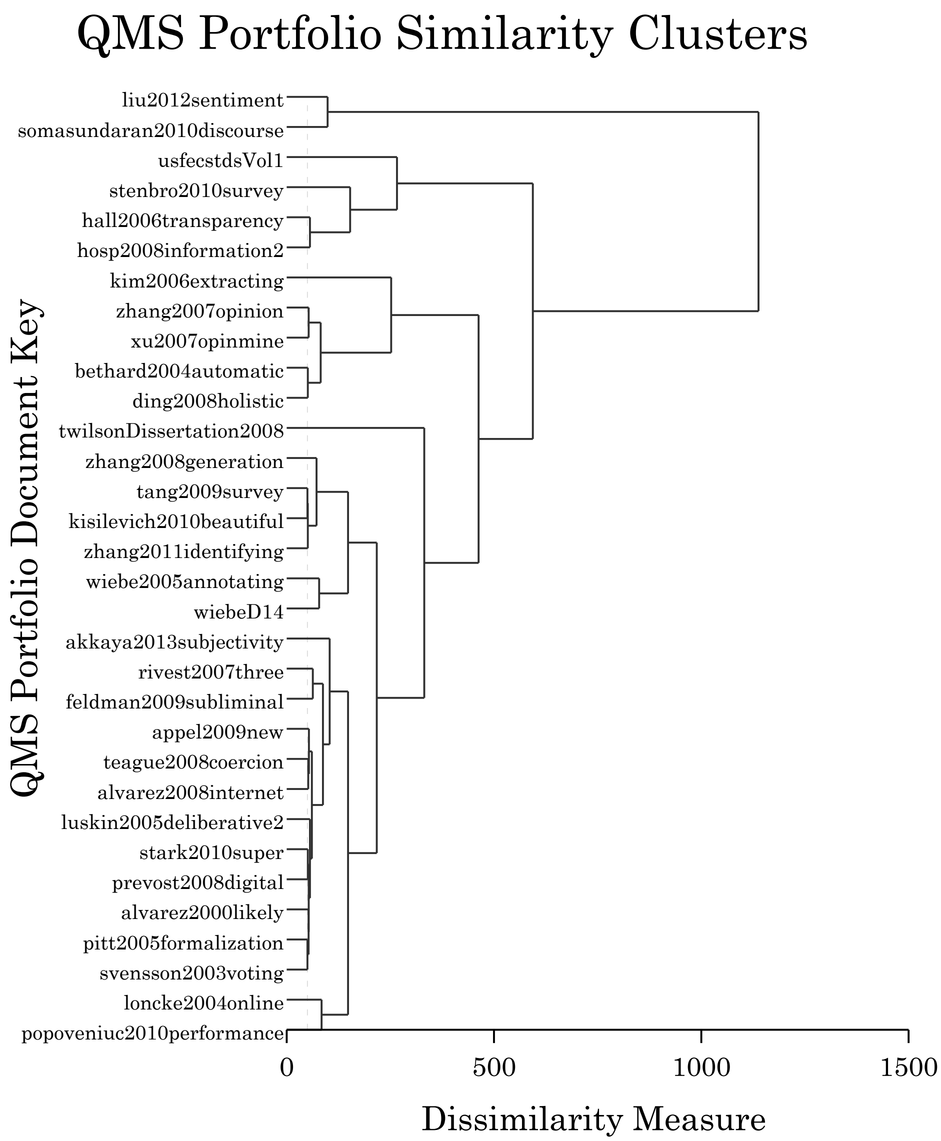


Figure 3.4: Dendritic graph showing relatedness of QMS portfolio documents based upon word frequencies from the text of Claim 3.3, "*Private states inform opinion expressions which may be counted by a voting system.*"

Figure 3.5: Word frequency map of Stenbro (2010).

a unique view of topic salience across the portfolio. The appendix enables a qualitative review of the treatment of the contents of the papers consumed through the QMS process.

The purpose of these word maps is to give a qualitative sense of the documents throughout the remainder of the QMS process. The creation and use of the word maps aligns with what Walsh & Downe (2005, p. 208) explains: "the study may be summarized loosely to draw inferred themes and concepts from the narrative."

The QMS process was augmented with aspects of explication (Chaffee, 1991). Chaffee (1991, p. 24) provides a definition: "nominal definition is an arbitrary name that lacks linking statements ... meaning analysis provides that kind of specification." This list of nominal concepts is meant to be brief—a starting point for exploring relationships amongst the concepts through meaning analysis in the next section.

An interesting outcome of the initial inspection of the various word maps, such as Figure 3.5, was the observation that very few verbs carry enough salience to be prominent in the maps. However, verbs define the interactions between and amongst concepts. Therefore, some additional effort was made to identify prominent verbs and include those in the nominal definitions.

Concepts relevant to *"private states."* Wiebe et al. (2005) is a widely cited paper on an ambitious corpus annotation project. In this important work, an annotation scheme for private states is proposed and then investigated. Wiebe et al. (2005, p. 167) describes privates states as, "*opinions*, *emotions*, *sentiments*, *speculations*, *evaluations*, and *other private states* . . . held by an *experiencer*, and optionally about a *target*."

Wiebe & Deng (2014, p. 1) investigates, "a deeper automatic interpretation of *subjective* language . . . [through] understanding implicatures [and] implicit sentiments (and *beliefs*)." Akkaya (2013) looked at using word-sense for disambiguation in support of subjectivity analysis. This work overlaps substantially with Wiebe & Deng (2014).

The definitional work in Wilson (2008) on private states is extensive and includes robust definitions for subjectivity, polarity, and attitude. Somasundaran (2010) includes an inventory of private state types—something not found elsewhere in the literature by this investigator. Table 3.2 summarizes the consolidated list of nominal concepts relating to private states.

Concepts relevant to opinion. Liu (2012) and Somasundaran (2010) are both comprehensive works on opinion mining and sentiment analysis. Therefore, the clustering algorithm grouped them at the top of the dendritic graph in Fig-

Table 3.2: Nominal concepts related to "private state"

attitude	opinion
belief	polarity
emotion	sentiment
evaluation	speculation
experiencer	subjectivity
holder	target
intention	

ure 3.4. Liu (2012) is a survey of recent literature which itemizes and defines the elements and methods used in opinion mining and sentiment analysis research.

In Liu (2012), definitions are provided for opinion, sentiment, emotion, aspect, and subjectivity. Various approaches to extracting these elements from free text are also reviewed. Liu (2012, p. 21) describes as, "[frameworks] to *transform* unstructured text to structured [opinion] data." Liu (2012) also refers to earlier work by the author where a canonical definition of opinion is developed as a quintuple of holder, entity, aspect, sentiment, and time.

In addition to the inventory of private states mentioned earlier, Somasundaran (2010) provides an excellent overview of opinion mining topics. This overview includes the influence of discourse relations on extraction of stance information using argument, polarity, and opinion frames.

The QMS portfolio includes nine (9) opinion identification studies. In some of the earliest work in opinion mining, Bethard et al. (2004) proposed the automated identification of opinion propositions, including the holder. Kim & Hovy (2006) developed a method of identifying an opinion, then connecting it to the holder and topic. Zhang et al. (2007) and Zhang & Ye (2008) connected relevance values for search results with opinion scores. Xu, Wong & Xia (2007) proposed OpinMine, a system for identifying opinionated sentences from a corpus, then extracting holder, polarity, and topic.

Table 3.3: Nominal concepts related to "opinion"

aggregation	polarity
ambiguity	proposition
argument	reason
aspect	relevance
domain	reporting
emotion	score
feature	sentiment
holder	structured data
noun-phrases	subjectivity
opinion	topic
opinion frame	transform
orientation	unstructured text
orientation strength	valuation

Ding et al. (2008) used a defined context of a product and given a set of features, and then classified reviewer comments according to polarity. Tang et al. (2009) extends opinion mining beyond the identification of the opinion, to methods of opinion aggregation and reporting. Additionally,Tang et al. (2009) includes a proposed method of identifying the reason for the opinion if available in the text. Kisilevich et al. (2010) uses online photo reviews to extend the polarity model of opinion valuation to be a real value. The algorithm used by Kisilevich et al. (2010) includes orientation, orientation strength, and an ambiguity value which reflects the amount of conflict between reviewers.

Opinion words and adverbial modifiers are not the only indicators of subjective speech. Zhang & Liu (2011) use opinion mining and sentiment analysis methods to identify positive and negative noun-phrases which represent opinions—typically given some domain knowledge.

Concepts relevant to voting systems. In the wake of the controversies surrounding the voting systems failures in Florida in the Presidential election in 2000, interest in voting systems scholarship expanded. Evidence for the increased rate of voting systems research is shown in Figure 3.1. It is not surprising, then, that the voting systems research included in the QMS portfolio happens to include research only from after the year 2000. NASED (2002) is a standards document published by the Federal Elections Commission, in the wake of the 2000 elections.

> "State and local officials today are confronted with increasingly complex voting system technology and an increased risk of voting system failure. Responding to calls for assistance from the states, the United States Congress authorized the Federal Election Commission (FEC) to develop voluntary national voting systems standards for computer-based systems." (p. 1-7)

Definition of voting system, accuracy, integrity, tallying, qualification testing, and auditing are all explained in some detail within NASED (2002). The core concepts which make up these definitions include ballot, mark, voter, vote, contests, candidates, issues, and data.

With the emergence of the Internet as a social platform after 2000, much of the research investigated the impacts of online voting. In early theoretical work on online elections, Loncke & Dumortier (2004) reviewed the legal challenges associated with online voting in elections—including integrity, verifiability, reliability, security, and audibility. Svensson & Leenes (2003) reviewed online voting activities in 13 countries, emphasizing effects on voter turnout. Hall (2006) reviews the history of transparency in voting systems in the United States and expounds the benefits of transparency of source code in voting systems software.

Stenbro (2010) is a survey of modern electronic voting system technologies used worldwide. It also includes a discussion of two electronic methods in development: Helios and E-Vote. The elements of analysis in this review include vote, voter, ballot, integrity, cast, election, tallying, auditing and encryption.

An extraordinary presentation of an algebra for the central principles of voting systems design is presented in Hosp & Vora (2008). These principles, each with an accompanying algorithm, are usability, integrity, privacy, verifiability, and robustness.

In this counter-technology paper, Rivest & Smith (2007) proposes three paper-based vote casting methods, with an emphasis on verifiability without encryption—each relying on some form of private or public receipt or acknowledgment. Rivest & Smith (2007) also reviews vote tallying protocols. Stark (2010) proposes a simple auditing procedure for votes recorded on manually submitted paper ballots.

Feldman & Benaloh (2009) demonstrates convincingly that ballot stuffing in encrypt-on-cast electronic voting systems can have coercive effects when incremental results are viewable by voters who have not voted yet.

Teague et al. (2008) also looks at coercion risk through simple progressive vote tallying. A legal challenge by voting systems scholars from Rutgers to New Jersey elections in 2004 is the subject of Appel et al. (2009). Faults identified by the study include usability, cartridge tampering, ballot corruption, and hardware problems.

In reviewing the effects of technology on elections, Alvarez & Nagler (2000), concluding that Internet voting is likely to introduce a material bias in election outcomes.

Later, Alvarez et al. (2008) reviews two elections (2005 and 2007) in Estonia conducted with the inclusion of votes from ballots submitted through Internet Voting. Prevost & Schaffner (2008) investigates the potential for bias in elections where Internet voting is used as a valid form of absentee ballot. Prevost & Schaffner (2008, p. 510) found that, "race and other socioeconomic factors do not affect the choice of Internet voting when it is used as an absentee voting method." Important work by Pitt et al. (2005) develops a voting protocol for multi-agent virtual organizations (MVO). The central elements in this protocol include agent, motion, ballot, voter, proposer, seconder, chair, and monitor.

Table 3.4: Nominal concepts related to "voting systems"

accuracy	reliability
auditability	robustness
ballot	security
candidate	tallying
casting	tampering
coercion	transparency
contests	turnout
data	usability
election	verifiability
encryption	vote
integrity	voter
peakedness	voter intentions
privacy	

In an important theoretical shift, Popoveniuc et al. (2010) challenges the emphasis on verifiability of voting systems by emphasizing models of verifiability of elections. An extended quote is included here because the thrust of this criticism is prescient for opinion mining and sentiment analysis scholarship in general, and this research, in particular.

> "focusing on end-to-end *verifiable elections* and not on voting systems: we care if the election outcome accurately reflects the *intentions of the voters*, regardless of whether the voting equipment is 'correct' or not. That is, it is ultimately the election that is checked, not just the equipment." (p. 1)

Lastly, Luskin & Fishkin (2005) make the case that if a deliberative process is integrated into a voting process, it can improve the quality of the decisions made. The concept of peakedness is described—an awareness of the issues which enables a population to overcome cycles of majority rule. The list of nominal concepts relating to a voting system is shown in Table 3.4. The next section consolidates the nominal concepts into a smaller listing of key concepts used in synthesis as part of the QMS process for theory development.

3.3 Analysis

The concept work was completed in Section 3.2. The analysis process can begin. The concepts from the literature relating to “private states” (Table 3.2), “opinion” (Table 3.3), and “voting systems” (Table 3.4), are conformed to the structure of the text of the original scoping statement in Claim 3.3.

The process of working through the classification of the concepts from the QMS portfolio is tedious. Though tedious, both the concept relations in Table 3.5 and the rationales which follow are essential to the next step in the QMS process. The result of QMS analysis is a table showing the relatedness of concepts from the QMS portfolio of private states, opinion mining and sentiment analysis, and voting systems literature, to the concepts of the QMS scoping assertion, Claim 3.3, “Private states inform opinion expressions which may be counted by a voting system.”

The purpose of this process is to coalesce the concepts from each aspect of Claim 3.3 in the hopes that further synthesis (shown in Section 3.4), can yield some useful theoretical propositions.

QMS advises practitioners to pursue an “interpretation of the current state of the art.” The heuristic used to align concepts from the literature with the concepts from Claim 3.3 is just that—an application of domain knowledge and peer review. The operative question for each concept taken from the literature, “to which concept in Claim 3.3 does this concept belong?” The resulting table of alignment between the concepts from the literature in the QMS portfolio and the concepts from Claim 3.3 is shown in Table 3.5.

In the following three sections, the concepts in each group of literature (“private state”, “opinion mining”, and “voting systems”) are discussed. Each concept is reviewed and categorized according to the concept classifications in Claim 3.3 (“private state”, “opinion expressions”, and “voting system.”)

Table 3.5: Alignment of private states, opinion expressions, and voting systems concepts with the text of Claim 3.3, *"Private states inform opinion expressions which may be counted by a voting system."*

	private states	**opinion expressions**	**voting system**
Concepts from "private states" literature.	attitude belief emotion intention opinion sentiment speculation† target		evaluation polarity sentiment‡ subjectivity
Concepts from "opinion mining" literature.	aspect domain emotion‡ feature opinion‡ proposition reason sentiment‡ topic	argument† noun-phrases† opinion frame† unstructured text	aggregation ambiguity orientation orientation strength polarity‡ relevance score structured data subjectivity‡ transform valuation
Concepts from "voting systems" literature.	candidate coercion contests voter intentions‡	ballot casting vote	accuracy auditability† data election encryption† integrity peakedness† privacy† reliability reporting robustness security† tallying tampering† transparency† turnout† usability verifiability

† - excluded as peripheral. ‡ - duplicate.

It is important to note that the emphasis is on the rationale for alignment to the original scoping statement in Claim 3.3. Definitional support is provided only as needed to explain the categorization. Where needed in the Section 3.4 (Synthesis) more detailed definitions are supplied.

Some of the concepts in Table 3.5 were eventually determined not to be relevant to the theoretical formulations which are described later in this paper. Those concepts are marked with a † and not discussed. Also, some concepts are represented in more than one (1) literature group. The analysis for purposes of alignment is the same regardless of the literature source, and so those concepts found in more than one (1) literature group are only discussed within the first group encountered. These duplicated concepts are marked with ‡ to indicate that the concept is discussed in a previous literature group.

3.3.1 Concepts from the QMS Portfolio Related to "Private States."

The rationale for categorizing each of the concepts in the private state literature within the QMS portfolio is discussed below.

Attitude. The concept of "attitude" directly relates to the concept of "private states." Wilson (2008, p. 1) describes attitude within the context of private states: "attributes of private states include ... the type(s) of attitude being expressed."

Belief. The concept of belief is a private state whose target is a proposition. As Wilson (2008, p. 1) explains, "a person may be observed to assert that God exists, but not to believe that God exists. Belief is in this sense 'private'."

Emotion. Emotions are informed by beliefs and desires and in turn inform intentions and behaviors. Thus, the concept from Claim 3.3 which "emotion" is most related to is "private state." Wiebe et al. (2005, p. 168) draws a more direct relationship, saying, “private state [is] a general term that covers opinions, beliefs, thoughts, feelings, emotions, goals, evaluations, and judgments.”

Intention. Wilson (2008, p. 117) defines intention as, “aims, goals, plans.” These types of intention represent pre-behavioral cognition, which is a “private state.”

Opinion. Perhaps the concept with the widest possible breadth of meaning and usage, is “opinion.” This present work has spent considerable energy establishing that “opinion” is a “private state.” The definition provided by Bethard et al. (2004, p. 1) shows this obvious relation: “a sentence, or part of a sentence, that would answer the question 'How does X feel about Y?'... not [including] statements verifiable by scientific data.”

Sentiment. The concept of “sentiment” has a dual nature in the literature—and is therefore shown in two columns in Table 3.5. On the one hand, “sentiment” is often defined as in the case of Somasundaran (2010, p. 46) : “emotions, evaluations, judgments, feelings and stances.” This definition aligns “sentiment” with “private states.” At other times, “sentiment” is used as a derived measure, as in Liu (2012, p. 19) where ”sentiment” is a value of: “positive, negative, or neutral, or expressed with different strength/intensity levels.” In this case, “sentiment” would be more closely associated with the counting function of the concept of “voting systems.”

Target. “Target” is the object associated with the “private state.”

Evaluation. The concept of “evaluation” from the private states literature has multiple uses. It most commonly refers to a researcher’s method of inference about the private state of a subject who has written some text. This usage aligns closely with the concept of a counting function in a "voting system" from Claim 3.3.

Polarity. The concept of "polarity" from the private state literature could be associated with the private state itself. At first it may appear that these are properties of the private state. However, in most of the literature, polarity is defined as Somasundaran (2010, p. 1) defines it: "whether the subjective expression is positive, negative or neutral." The concept of "polarity" relies on a scale of classification whose value is derived by inference, rather than being an inherent property of a private state. Put another way, "negative" is not an attitude, emotion, or belief—it is a valuation. The concept of valuation aligns closely with the concept of counting within a "voting system", so "polarity" is classified as such.

Subjectivity. "Subjectivity" (or "subjectiveness") is by definition derived from the existence of a private state. Therefore, in the context of private state scholarship, subjectivity is a value derived by inference—i.e., whether or not a particular statement is subjective. Somasundaran (2010, p. 1) defined subjective statements as, "expression of speculations, evaluations, sentiments, beliefs, etc. (i.e., private states)."

Wilson (2008, p. 2) said a subjective statement, "contains one or more private state expressions." The opinion mining literature concurs with this conceptualization of subjectivity. Liu (2012, p. 27) summarizes the types of subjective expressions, "opinions, allegations, desires, beliefs, suspicions, and speculations."

The concept of "polarity" (as a measure) is closely related to the concept of "voting systems". As a measure, the concept of "subjectivity" is also a derived value. Therefore, it is categorized as a concept relating to "voting systems."

3.3.2 Concepts from the QMS Portfolio Related to "Opinion Mining."

As shown in Table 3.5, opinion mining literature includes concepts related to all aspects of Claim 3.3. A number of these concepts are not obviously categorized, and merit extended explanation.

Components of Opinion. The concepts of "aspect", "domain", "feature", "sentiment", and "topic", from the literature on opinion mining are all components of an opinion. "Opinion" itself has been shown to be a "private state"; therefore, all of these concepts align closely with the concept of "private states" from Claim 3.3.

Proposition. The concept of a "proposition" has many applications and usages, but the particular meaning within the QMS portfolio is that of a "stated belief" about some event or phenomena. The usage is often connected with converting propositional statements into some form of structured opinion. Propositions are a linguistic construct which carried belief. Wiebe & Deng (2014, p. 13) explains, "Events, on the other hand, are not themselves propositions ... [propositions require that] the source has to believe something about the event." "Belief" is an important concept in understanding private states, so "proposition" is classified as belonging to the "private states" concept of Claim 3.3.

Reason. "Reason" refers to "opinion-reason mining" in the context of Tang et al. (2009). Tang et al. (2009) is an oft-cited paper within the QMS portfolio because it goes beyond focusing on the opinion itself. While the lexical manifestation of the "reason" is some text saying what the "reason" is, the "reason" itself is rooted in the private state space of the person expressing the opinion. If the opinion itself is a private state, then the forces which motivate the behavior to express an opinion in a certain way are also a private state. Therefore, “reason” is a private state concept.

Unstructured Text. Within opinion mining and sentiment analysis, the concept of “unstructured text” can be a little misleading. After all, if the text is unstructured, then how can we understand what it says? The criteria for unstructuredness, however, in opinion mining research is whether or not the opinion expressions are explicitly and unambiguously available. If the elements of the operational definition of opinion from Liu (2012), above, are all identifiable with the text of the opinion expression, then the text has structure. In all other cases, as it relates to opinion mining and sentiment analysis, the expression of opinion is “unstructured text.” Therefore, “unstructured text” is a concept relating to opinion expressions.

Aggregation. The concept of “aggregation” from the opinion mining literature, can be easily categorized as a concept relating to a system of counting opinions, or the concept of “voting system” from Claim 3.3.

Ambiguity. “Ambiguity” is a property of the expression of opinion, and so is most closely related to the concept of “opinion expressions.”

Orientation. ”Orientation” and ”orientation strength”, like “polarity” as discussed above, are values derived by inference from the unstructured text through algorithmic means, and so are categorized within the “voting systems” concept of Claim 3.3.

Evaluation (score, valuation, and relevance). Lastly, “score”, “valuation”, and “relevance” are concepts which relate to a post-inferential processing of structured representations of opinion. These are concepts closely connected with concepts such as “opinion search”, “opinion quality”, and “reputation.”

Transform. The concept of ”transform” in the opinion mining literature embodies the process of transforming “unstructured text” to ”structured data”. Therefore, both “transform” and “structured data” are concepts relating to the “voting systems” concept of Claim 3.3.

3.3.3 Concepts from the QMS Portfolio Related to “Voting Systems.”

Candidate and Contest. In simplest terms, “candidate” is a possible choice, and “contest” is a question. In the literature, “candidate” is discussed primarily in terms of its physical manifestation, i.e., a person (Rivest & Smith, 2007; Appel et al., 2009; Popoveniuc et al., 2010). However, for purposes of this analysis and alignment of the concepts within Claim 3.3, it is an individual’s understanding of this concept which is highly relevant. In this sense, the “candidate” represents a specialized type of “target” discussed previously as an aspect of a “private state.” “Contest” represents a propositional space into which the “candidate” fits. Therefore, both “candidate” and “contest” are most closely related to “private state.”

Coercion. “Coercion” refers to undue influence of one party on the intentions of another. The term “undue” indicates not due, i.e., outside of the establish protocols governing those intentions. Teague et al. (2008, p. 1) connects coercion with voting systems this way: “it is important that voters are free of coercion, votes are tallied correctly and this is seen to be the case . . . [and] voting systems must take particular care to prevent a voter from being able to prove to a coercer how they voted.” Teague et al. (2008) investigates the scenarios in which electronic voting systems enable or even encourage coercion within elections. “Coercion” is an environmental factor. It may affect a person's intentions. Therefore, "coercion" is most closely related to the concept of a "private state."

Ballot. A "ballot" is a physical or electronic representation of topic or question. Therefore, it is most closely related to the concept of "opinion expression."

Cast. The concept of "cast" or "casting" from the voting systems literature, overlaps substantially with the concept of "opinion expressions" from Claim 3.3. "Cast" represents the behavior associated with transforming a private intention into a public expression of opinion.

Vote. As shown previously in Definition 3.1, a vote is an expression of an opinion, and so is aligned with "opinion expression."

Data. "Data", in this context, represents the record of events associated with the operation of a voting system. This record of events, in the case of ballots and votes, is closely related to the concept of "structured opinion." NASED (2002, p. 2-46) connects voting systems, votes, and data, as follows: "All [voting] systems shall provide a means to consolidate vote data from all polling places."

As shown in the United States Presidential Election of 2000, a ballot does not always equate to a precise expression of opinion. Some processing by a system (human or electronic) is required to establish the "data" associated with the ballot. Therefore, the concept of "data" is most closely related to "voting system."

Aspects of Voting Systems. As shown previously, concepts which are properties or components of a "private state", such as "target", are themselves most closely related to the concept of a "private state." The following concepts are ways of measuring the performance characteristics of “voting systems”: “accuracy”, “integrity”, “reliability”, “verifiability”, “usability.”

The precise definitions of these concepts turn out to be important to the subsequent synthesis in Section 3.4. However, for purposes of alignment, it is sufficient to observe simply that these concepts most closely align with the concept of “voting systems.”

Summarization. The concepts of “election”, “reporting”, and “tallying”, are all indicative of post-processing of “structured opinion” done by a “voting system.” Therefore, these concepts are aligned with the concept of “voting system.” “Election” is the decision-making process which acts upon summarizations of “structured opinion.” “Tallying” and “reporting” are also discussed in the next section.

3.4 Synthesis

It is important to express that while this paper reads in serial fashion, the execution of the research was highly iterative—though always operating within the scope defined in Section 3.2.3. The following nominal definitions were selected as being representative of others in the portfolio, based upon the strength of the population process described above. The justification is that these documents are given to be highly relevant to the inquiry at hand by virtue of the selection process used. Kim & Hovy (2006, p. 5) states, “Despite the lack of a precise definition of sentiment or subjectivity, headway has been made in matching human judgments by automatic means.”

3.4.1 Informal Theoretical Structure

The following is a listing, with brief explanations of the actions and relations shown on Figure 3.6. The numbers below correspond to the labels in Figure 3.6.

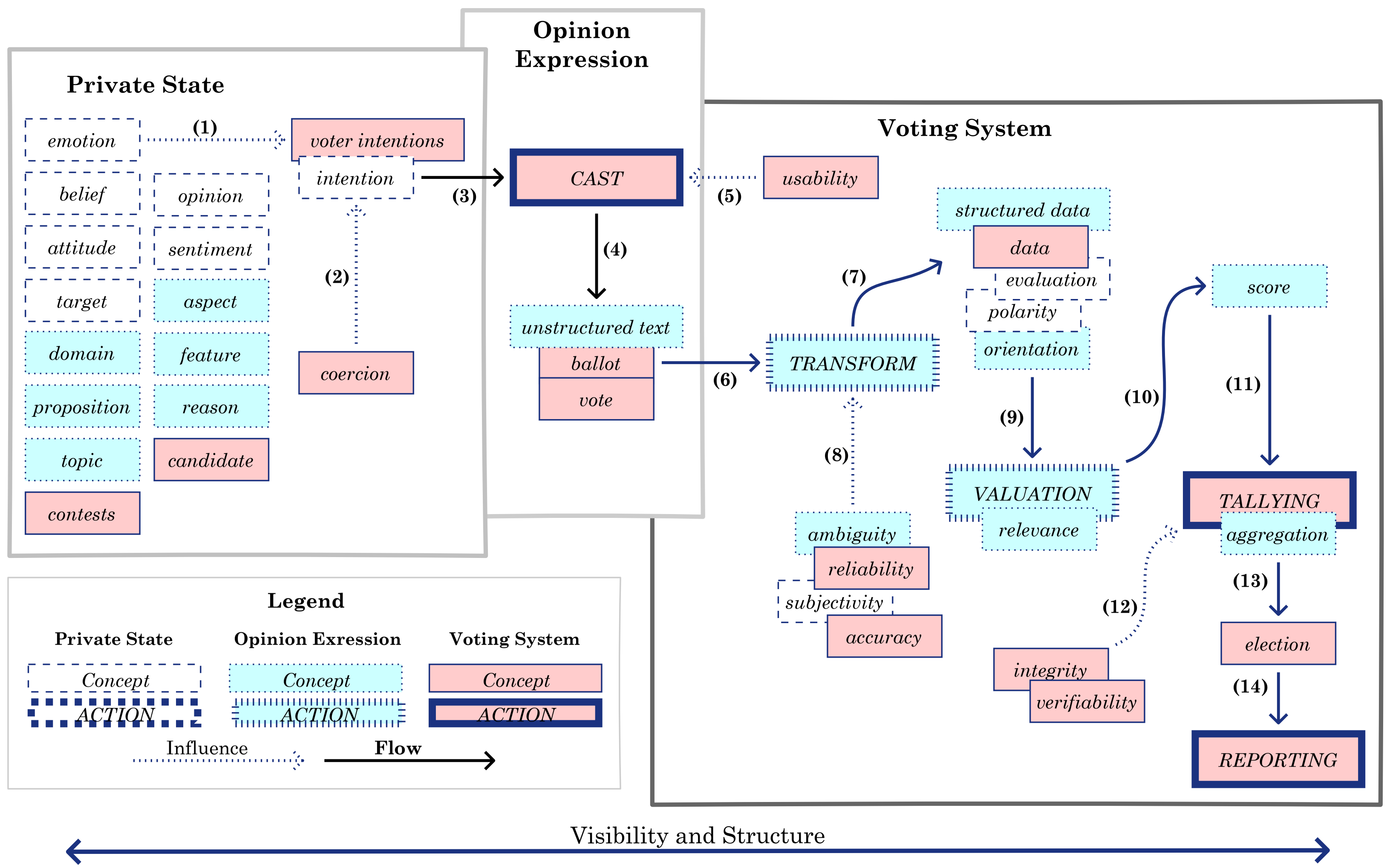


Figure 3.6: Synthesis of aligned QMS portfolio concepts into initial theoretical model using a continuum of visibility and structure. Model shows constructs, actions, influences, and flows.

1. Emotions, attitudes, and beliefs influence intention.

2. External factors influence intention.

3. The behaviors of casting a ballot containing one or more votes, or expressing an opinion in free text, are non-private persistent projections of private intentions.

4. Regardless of the medium of expression, there are environmental forces exerted which influence the projection of intention to an expression of opinion. Sometimes these forces are coercive, other times merely informative.

5. Research in opinion mining and sentiment analysis revolves largely around the transformation of unstructured representations of opinion into structured representations. In a voting systems context, the opinion expression is a ballot, either electronic or paper, but the result is the same: a structured representation of the expressed intentions.

6. The transformation in (5), is influenced by characteristics of the expressed opinion. Ambiguity and subjectiveness are significant factors in the ability of any transformation process to reliably and accurately convert a persisted expression of an opinion in any form into a structured representation of opinion, as data.

7. The evaluation of the data derived from (5) for a particular expression of opinion often includes an assessment of the veracity or weight of that data. The result is something akin to a weighted opinion, or score. For reputation systems, the weight associated with one person's expression of opinion might be higher than another's. In the case of a formal election, the valuation process might include a determination that the data indicates fraud. In that case, the ballot might be given zero weight and scored as aberrant, to be excluded from tallying (8).

8. Tallying takes the scores from (7) and counts them to produce an aggregation of the scores.

9. The perfect tallying process has 100% integrity and 100% verifiability (Hosp & Vora, 2008). Integrity refers to ensuring that all the information coming out of (7) is included in the aggregation resulting from (8). Verifiability is the ability of a system to prove its integrity. The integrity and verifiability controls associated with a voting system will have a direct influence on the outcomes from tallying (8).

10. Reporting is the process of transforming the aggregations from tallying (8) and making them available for consumption. In the case of formal elections,

> that consumption is an electoral board which must certify the reported outcomes. In the case of opinion in social media, that reporting process might include a geographic representation of votes cast online.

3.4.2 Formal Theoretical Structure

The process of formalizing the elements of our theoretical model requires that authoritative works outside of the QMS portfolio be consulted. The purpose of these sources is not to change the alignment or informal structure, but to anchor the finished product in the best possible related scholarship. The name for the formal theoretical model developed through this research is called the States, Processes, Effects, and Quality (SPEQ) Model of Opinion Mining and Sentiment Analysis. The complete SPEQ model is shown in Figure 3.7, including linkages to the literature and informal theoretical model in Figure 3.6.

3.4.3 States, Processes, Effects, and Quality (SPEQ)

As shown in Figure 3.7, SPEQ consists of six processes, seven opinion states, and five quality measures (which includes three integrity measures). The SPEQ processes are what govern the progression of any opinion through the seven (7) states. The quality measures reflect the amount of "bias" or "error" present in any opinion state transition. For purposes of this discussion, "bias" can be thought of as systemic error in the direction of a particular outcome or class of outcomes—error not due to the nature of the input. "Error," is the deviation from the original input which can be attributed to how the process operates on that specific input. A perfect process operates on the input and produces the corresponding output with zero "bias", and zero "error". The SPEQ states and quality measures account for the "chain of custody" of affect from origination by belief and desire through opinion mining and reporting. The following is a review of the six (6) processes. Because

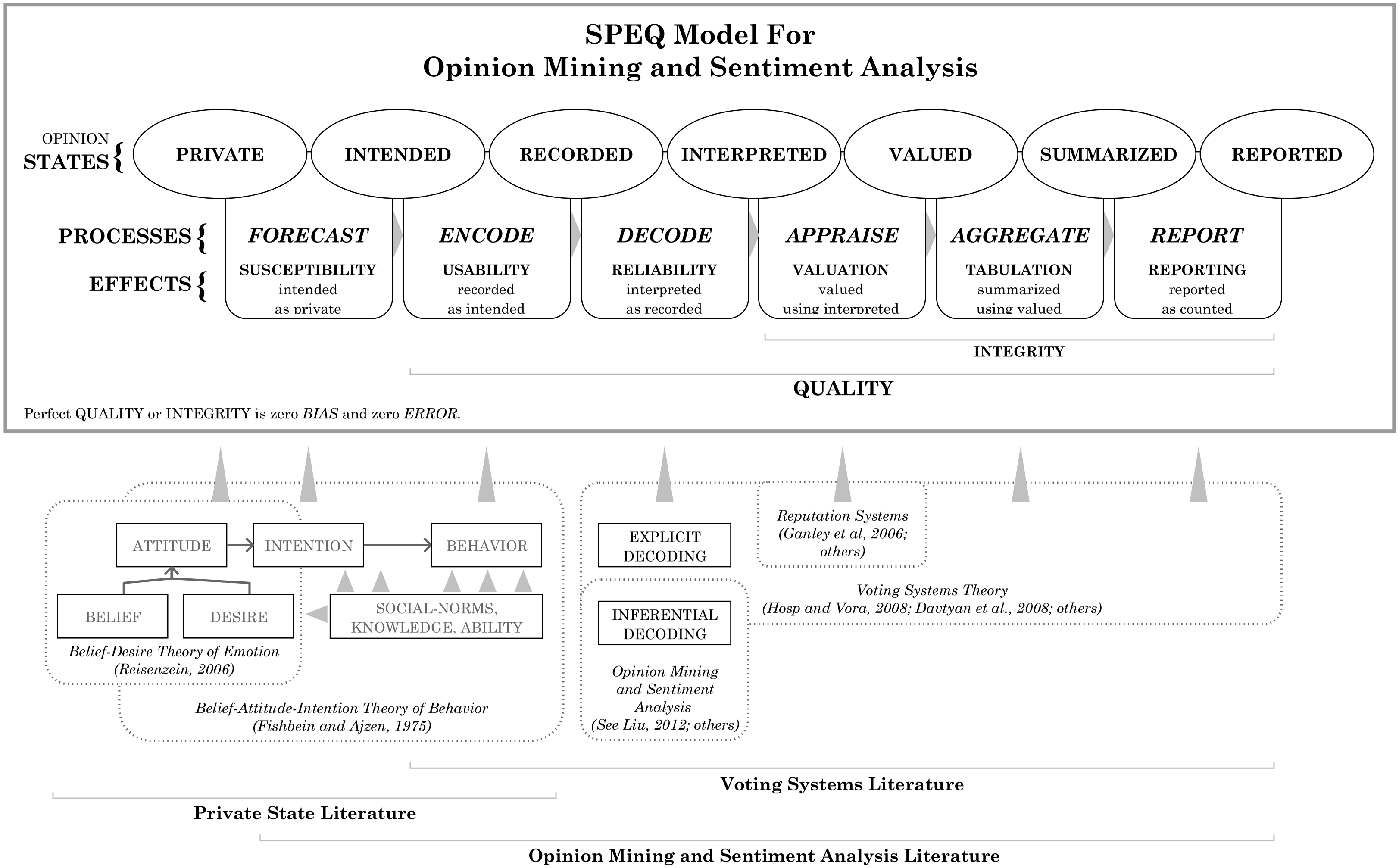


Figure 3.7: The States, Processes, Effects, and Quality (SPEQ) model for opinion mining and sentiment analysis.

the theoretical framework is new, this review will be a thematic review, rather than a technical review since much work has to be done to formalize these elements.

Forecast: *Private* Opinion to *Intended* Opinion. The originating state of an opinion in SPEQ is "private." As the name implies, a "private opinion" is known only to the holder and is based on the holder's beliefs and desires (Reisenzein, 2006).

The second state of opinion in SPEQ is "intended." The "intended state" is the state which resolves the conflict (if any) between the holder's "private opinion" and the holder's intentions regarding future impacts of his or her behaviors. These behaviors include expressing or acting upon the "private opinion." Those intentions may be influenced by external factors such as social norms (Fishbein & Ajzen, 1975). Therefore, in SPEQ, "forecast" translates a "private opinion" into an "intended opinion."

In the example shown in Figure 3.8, it is shown how the opinion of a person, Vivian, progresses through the SPEQ model. Vivian's "private opinion" equates to "I feel." Her "private opinion" is, "I like Smith as a candidate"; however, Vivian is subject to external forces and considerations, which influence her intentions. Vivian will "forecast" whether or not there is anything in her environment now or in the future which would cause her to form an intention different from her "private opinion." Vivian appears to be susceptible to external considerations or forces, and the outcome of her forecast is an intended opinion of, "I intend to vote for Jones."

Encode*:* *Intended* Opinion to *Recorded* Opinion. If a person wants to make their intentions known to others, then that person will need to encode their "intended opinion" in some way to get it beyond the boundaries of their awareness.

The method of encoding can be a gesture, speaking, writing, posting to a blog or using a voting machine.

The outcome of "encode" is a recorded projection of the person's "intended opinion." In the case of a gesture, that projection could include raising a hand or nodding a head or a thumbs up, recorded by the individuals seeing the gesture. In social media, "encode" can take the form of a micro-blog post or text message, or Facebook posting.

In the pathological example shown in Figure 3.8, encode equates to "I expressed" my opinion. Vivian's "intended opinion" is "I intend to vote for Jones." However, there are some usability problems (Hosp & Vora, 2008) with the mediating device she uses to encode her "intended opinion." As a result, rather than a vote for "Jones", Vivian's "recorded opinion" is a vote for "Janes." Apparently, she misunderstood some aspect of the ballot.

Extending the example to social media, Vivian also intends to create a wave of support for Jones on Twitter. Because Vivian is a Jones supporter, she "intends" to post to her Twitter feed, #jones+++. However, the type-completion on Vivian's phone produces #junes+++, and because Vivian's phone is hard to read in sunlight, she does not catch the error until after she transmits to her followers.

Decode: *Recorded* Opinion to *Interpreted* Opinion. Encoded, or expressed, opinions are typically intended to be decoded, or interpreted. In SPEQ, "decode" takes a "recorded opinion" and applies a decoding process to produce an interpretation of that "recorded opinion." This interpretation is the "interpreted opinion", which in opinion mining and sentiment analysis research is often a structured representation of the original "recorded opinion", or data. Therefore, this SPEQ process is central to the issues raised by this paper for opinion mining and sentiment analysis research.

In Figure 3.8, “interpreted opinion” equates to “I am understood.” While Vivian’s recorded opinion was a vote for “Janes”, there was a problem with the chads on her ballot. The ballot reader produced an “interpreted opinion” of “Vivian votes for Dewey.” This type of error is highly relevant to opinion mining and sentiment analysis research, so it is worth examining in light of an example.

Here is an example “recorded opinion” from a Twitter corpus, used in a number of studies, including Bizau et al. (2011):

Example 3.1. @Scordellis1 I haven’t seen inception yet. I wana c toystory

What is the perfect “interpreted opinion” for the target “Inception” in Example 3.1? Does this person feel positive? What does he or she mean by “yet”? It may be a propositional opinion (Bethard et al., 2004) with positive orientation. Alternatively, it may be a propositional opinion with a negative orientation—by virtue of the statement expressing an interest in seeing Toy Story. In Bizau et al. (2011), Example 3.1 is included in the test set for negative sentiment toward the movie “Inception.”

The quality measure for “decode” is “reliability”, or “interpreted as recorded.” If a decoding mechanism is perfectly reliable, then there is zero error, and zero bias—i.e., an exact correspondence between the meaning of “recorded opinion” and the meaning of “interpreted opinion.” This difficulty highlights the central challenges of opinion mining and sentiment analysis research, and how SPEQ can help clarify both the phenomena being studied, and the quality of the findings.

Appraise: *Interpreted* Opinion to *Valued* Opinion. The “appraise” process is derived from the “valuation” concept discussed Section 3.3 and shown in Figure 3.6.

“Appraise” takes an “interpreted opinion” and scales its value according to whatever value system is in effect, producing a “valued opinion.” For most opinion counting applications, the "interpreted opinion" is the same as the "valued opinion." That is, there is a 1 to 1 correspondence between a particular "interpreted opinion" and its influence on a subsequent aggregation. In some cases, however, not all votes are valued equally. Some reputation systems might use voter karma (Ganley & Lampe, 2006) to weight the votes of one user more heavily than the vote of another.

Aggregate: *Valued* Opinion to *Summarized* Opinion. In SPEQ, the process of "aggregate" takes a "valued opinion" and includes it with other opinions for purposes of seeing what the totals are.

If there is perfect "tabulation integrity" (zero bias and zero error) then the resulting "summarized opinion" is a perfect representation of the relevant population of "valued opinion(s)."

In the hypothetical example in Figure 3.8, "aggregate" equates to "I am counted." The "valued opinion" is a vote for "Dewey" (the "interpreted opinion" is unchanged by "appraisal"). However, through a tabulation error, all "Dewey" votes were counted for "Johnson." Therefore, the summarized outcome is "Johnson wins!"

Report: *Summarized* Opinion to *Reported* Opinion. Lastly, "summarized opinion" is most helpful if that summarization is made available to inform interested persons. This process is called "report" in SPEQ. "Report" transforms the "summarized opinion" into a "reported opinion" consumable by those interested persons. The integrity measure for "report" is whether "summarized opinion" and the "reported opinion" include the same meaning.

Wrapping up the pathological example in Figure 3.8, "report" equates to "I am informed." The "summarized opinion" shows "Johnson wins!" However, Vivian, who is a mild-mannered voter by day is a malicious hacker by night. She had modified the voting system report generation program code to produce a report showing "Smith wins!" regardless of the "summarized opinion."

3.4.4 SPEQ Quality: Bias and Error

SPEQ provides a more granular and specific framework for opinion mining and sentiment analysis. One element of SPEQ, which might be particularly useful for opinion mining and sentiment analysis, is the need to account for "bias", and "error" when reporting findings. In conventional opinion mining and sentiment analysis studies, the "gold standard" is often the outcome of an arbitration process amongst a group of annotators. In the discussion of the results of the study, the authors might suggest, "an accuracy value of 94% was achieved." However, what does this really mean? For example, Ku et al. (2006) uses annotators to assign polarity values to words, sentences, and documents. The inter-annotator agreement was 68% on average at all three levels, for three annotators. The resulting arbitrated polarity assignments became the "gold standard" of the study. A precision value of approximately 61% was then achieved after applying an opinion mining algorithm to the corpus. Here are the findings from Ku et al. (2006):

> "Utilizing the sentiment words mined together with topical words, we achieve f-measure 62.16% at the sentence level and 74.37% at the document level. Involving topical words enhances the performance of opinion extraction." (p. 8)

SPEQ changes the rhetoric around the outcomes of opinion mining and sentiment analysis, by increasing the granularity of analysis. In the case of Ku et al. (2006), an application of SPEQ may have yielded the following explanation.

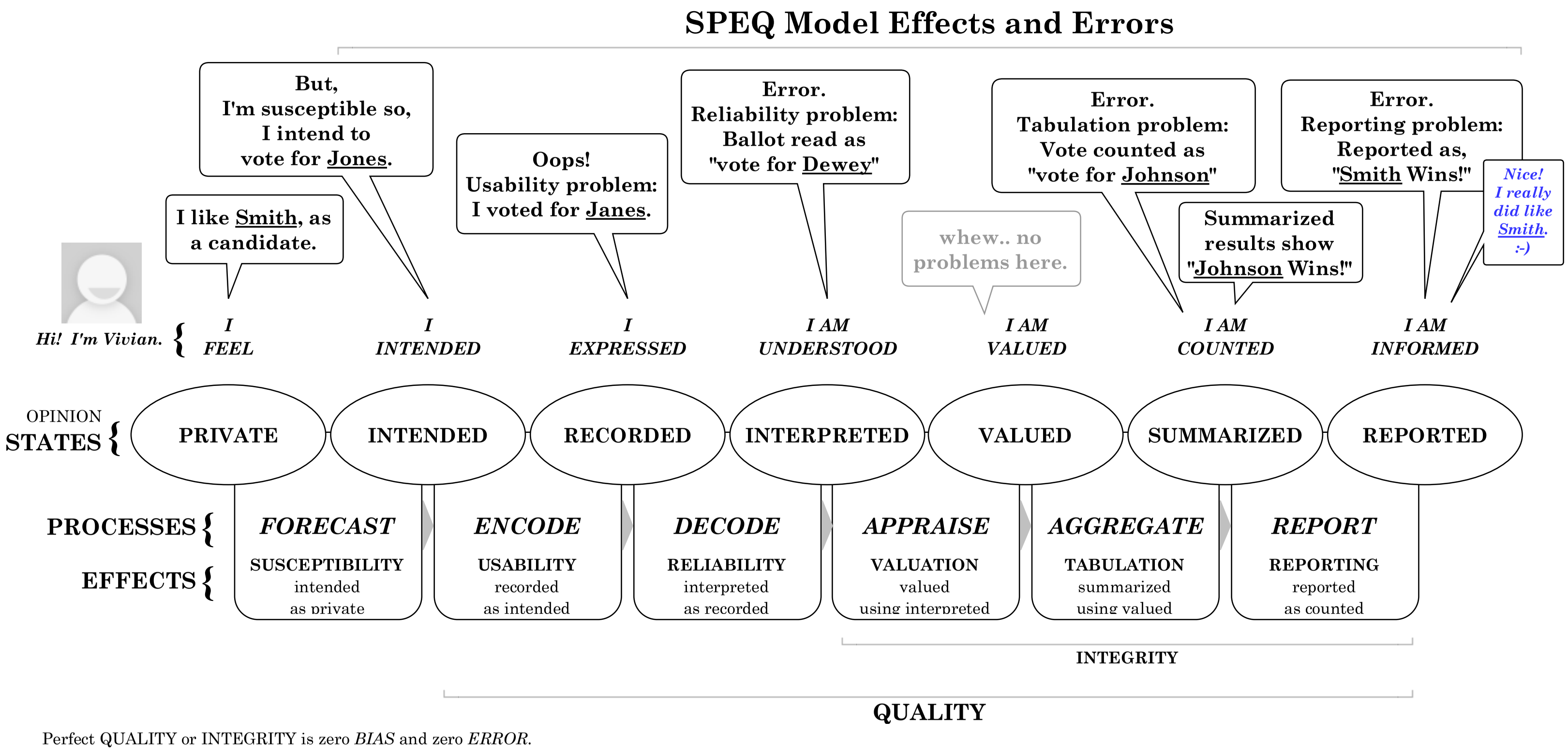


Figure 3.8: SPEQ example use case with effects and errors

"Two decoding mechanisms are compared in this research: an annotator panel (PANEL), and our algorithm (ALG). The recorded opinions were decoded using both PANEL and ALG.

The SPEQ error and bias values for PANEL are estimated to be e_{panel} and b_{panel}; the estimated error and bias values for ALG are e_{alg} and b_{alg}.

When comparing interpreted opinions of of PANEL and ALG, an F-measure of X was achieved utilizing the sentiment words mined together with topical words. Adjusting for the SPEQ quality of PANEL and the quality of ALG, the f-measure used to evaluate the null hypothesis is Z."

3.4.5 The Missing Opinion-related Verb: Voot?

To this point in opinion mining and sentiment analysis scholarship, the naming of key concepts in the literature has been muddled. An extended quote is provided below to amplify the issue as explained by Liu (2012):

> "There are also many names and slightly different tasks, e.g., sentiment analysis, opinion mining, opinion extraction, sentiment mining, subjectivity analysis, affect analysis, emotion analysis, review mining, etc. However, they are now all under the umbrella of sentiment analysis or opinion mining . . . we will use the term opinion to denote opinion, sentiment, evaluation, appraisal, attitude, and emotion. However, these concepts are not equivalent. We will distinguish them when needed. The meaning of opinion itself is still very broad. Sentiment analysis and opinion mining mainly focuses on opinions which express or imply positive or negative sentiments." (p. 7)

It is hoped that an outcome of this paper is that the concept of "opinion mining" itself is shown to be inadequate. Opinions are private states and cannot be "mined." SPEQ shows that opinions have multiple states, processes which govern the translation, and effects which introduce bias and error in those processes. While SPEQ itself clarifies the problem, there is still a gap in the current vocabulary of opinion expression. As evident in the QMS alignment work above, a word, specifically a verb, is missing. "Cast" is what someone does to encode their "intended opinion" so as to be decoded by a voting system.

It is interesting to note in Figure 3.6 that there is no verb from the opinion mining and sentiment analysis literature for the encoding of an "intended opinion" in a less formal setting for decoding. Such a setting would include social media. It would be convenient for many purposes if there were a word which carried the concept of an encoded (or formal) opinion expressed in an informal setting. This "thing" is what all the opinion mining and sentiment analysis is processing every day—but it has no name.

Such a proposition is also motivated by the introduction of SPEQ. SPEQ is a big step forward. It provides an integrated model covering private state, opinion mining and sentiment analysis and voting systems concepts. However, the formalisms which could develop around SPEQ would likely re-use the existing term, "vote" in informal setting. Alternatively, a prefix such as "social" could be added to get "social vote", but then the prefix muddies the waters as badly or worse. Perhaps it is worth introducing a new word to carry the meaning, to keep the distinctions clearer.

So, in the continuing spirit of the appeal for "daring generality" (Albig, 1957), this investigator proposes the introduction of the word "voot" to fill the void.

"Voot" is a Dr. Suess-like combination of the word "vote" and the word "cartoon." The semantic conjunction "voot" carries the desired semantics of a formally expressed but informally recorded vote.

The use of "voot" allows for a formal study of an informal social construct, without resorting to the word "vote." Without "voot" (or some other equivalent word), the concept of "formally expressed (i.e., I want to be counted) but informally recorded opinion" has no grounding lexeme.

By way of application, then, “recorded opinion” from an official or formal source becomes a “vote.” When decoded into an “interpreted opinion” from an informal source, the ”recorded opinion” becomes a “voot.”

So, “get out and voot” has a very specific meaning. It means, go out and record your opinions through informal channels, such as social media—and do so in a way that you can expect to be counted. A potential “voter” expressing herself on a blog, is a “vooter” who is “vooting.”

3.4.6 Vooting Systems Research

Another benefit of the SPEQ model is the formalization of the study of the characteristics of opinion mining and sentiment analysis research itself. One of the lessons learned from voting systems scholarship is that “quality” and “integrity” can be difficult concepts to precisely defined—and very difficult concepts to deal with operationally.

Opinion mining and sentiment analysis may benefit from spawning a discipline around the formal measurement of the systems they develop. These opinion mining and sentiment analysis systems turn “recorded opinions” in informal and unstructured representations into “voots.”

Thus, a new field of exploration is defined, organized around SPEQ, which can only help to improve the granularity of opinion mining and sentiment analysis research. Each state, each process, each quality measure, may be a field of study unto itself.

3.5 Conclusions

It has been shown in this paper that there is a close semantic relationship between the word “opinion” and the word “vote.” It has also been shown that while opinion mining and sentiment analysis scholarship has been propelled to expansive growth with the proliferation of social media, voting systems scholarship has stagnated in recent years. Moreover, it has been shown that the explosion of empirical work on methods of opinion mining and sentiment analysis has left key concepts within the field under-defined, and unanchored to rich theoretical models. This paper was motivated by the observation that while the two focal concepts of "vote" and "opinion" are closely linked, the two research disciplines are not linked theoretically, operationally, or in scholarly production in the literature.

An exhaustive (and, allowing personal reflection, tedious) Qualitative-Meta-Synthesis (QMS) analysis of literature from private state, opinion mining and sentiment analysis, and voting systems research showed conceptual overlaps among many concepts. These overlapping relationships enabled the development of a set of primitives and eventually the States, Processes, Effects, and Quality (SPEQ) model for opinion mining and sentiment analysis.

The consequences of SPEQ are potentially far reaching. SPEQ provides a theoretical and operational perspective which clarifies core elements of opinion mining and sentiment analysis. Currently, these exist in a murky and isolated pool of overlapping definitions and algorithms not grounded in any formal measure of quality to guide interpretation of findings.

With SPEQ, inquiries into the field of opinion mining and sentiment analysis can leverage a rich framework of seven distinct states of opinion, six processes which govern transitions between those states, and five measures of quality (including 3 measures of integrity). SPEQ was developed leveraging scholarship across social psychology, opinion mining and sentiment analysis, and voting systems, and so defines an end-to-end model of the "chain of custody" of opinion.

The QMS process followed in this research also demonstrated that language itself has not kept pace with the "things that happen" in the wake of social media proliferation. The term "voot" was proposed to represent a "vote" expressed in an informal and unofficial setting or medium. The use of the term "voot" may provide an important and meaningful semantic anchor for a large class of actions, concepts, and relationships for which today even academics appear uncertain.

Lastly, useful conclusion from the development and definition of SPEQ is that new avenues of approach for opinion mining and sentiment analysis are available. SPEQ provides a granularity of analysis which may enable more specific types of research with more clearly stated and meaningful outcomes.

Limitations

While SPEQ is an innovation introduced in this paper, the theoretical limitations seem difficult to pinpoint. This statement regarding the veracity of SPEQ may seem paradoxical, especially because SPEQ introduces new constructs. However, challenges to SPEQ would have to come from the literature—and the literature was the source for the development of SPEQ. The final QMS portfolio included 1600 pages of related scholarly works. These were rationalized through an intensely quantitive process and analyzed both through human cogitation and sophisticated linguistic analytical tools (CiteScan.)

Rather than with SPEQ itself, the limitations of this research are the speculative rationales used at many points in the QMS process. Liberties were certainly taken with definitions of key terms for which large bodies of scholarship exist. There are almost certainly large and conspicuous gaps in the literature considered. It may be that a similar process with a more thoughtful consideration of the literature and relevant concepts might produce a richer and more comprehensive model.

Also apparent, is the lack of mathematical proofs which govern the nature and implementation of "bias" and "error." A fair amount of hand-waving at these important aspects highlights a rich area of future definitional work.

Recommendations

It is the hope of this investigator that SPEQ will motivate other researchers to think more holistically about vooting systems.

SPEQ presents a tremendous opportunity for researchers in the field of vooting systems research. By leveraging—or even criticizing and extending SPEQ, it may be possible to help the field produce more rigorously defined, more complete, and more useful theoretical and operational models.

Even a debate around the efficacy of SPEQ would be a tremendous victory for the field. Opinion mining and sentiment analysis researchers would spend a little more time on vooting systems theory.

Coincident with a raging theoretical debate about the merits of SPEQ, a fruitful area of inquiry would be to formalize the nature of "bias" and "error" in each of the six processes within SPEQ. Such a formalization would help all empirical works underway migrate toward a common representation of their methods, concerns, risks, results, and recommendations.

BIBLIOGRAPHY

Akkaya, C. (2013). *SUBJECTIVITY WORD SENSE DISAMBIGUATION: A TOOL FOR SENSE-AWARE SUBJECTIVITY ANALYSIS*. PhD thesis, University of Pittsburgh.

Albig, W. (1957). Two decades of opinion study: 1936-1956. *Public Opinion Quarterly*, *21*(1), 14.

Alvarez, M., Hall, T., & Treschsel, A. (2008). Internet voting in estonia. Technical report, VTP Working Paper.

Alvarez, R. M. & Nagler, J. (2000). Likely consequences of internet voting for political representation, the. *Loy. LAL Rev.*, *34*, 1115.

Appel, A., Ginsburg, M., Hursti, H., Kernighan, B., Richards, C., Tan, G., & Venetis, P. (2009). The New Jersey voting-machine lawsuit and the AVC Advantage DRE voting machine. In *Proceedings of the 2009 conference on Electronic voting technology/workshop on trustworthy elections*, (pp. 5–5). USENIX Association.

Bethard, S., Yu, H., Thornton, A., Hatzivassiloglou, V., & Jurafsky, D. (2004). Automatic extraction of opinion propositions and their holders. In *2004 AAAI Spring Symposium on Exploring Attitude and Affect in Text*, (pp. 2224).

Bizau, A., Rusu, D., & Mladeni, D. (2011). Expressing opinion diversity.

Chaffee, S. H. (1991). *Explication*, volume 1. Sage Publications, Incorporated.

Ding, X., Liu, B., & Yu, P. S. (2008). A holistic lexicon-based approach to opinion mining. In *Proceedings of the 2008 International Conference on Web Search and Data Mining*, (pp. 231–240). ACM.

Feldman, A. & Benaloh, J. (2009). On subliminal channels in encrypt-on-cast voting systems. In *Proceedings of the 2009 conference on Electronic voting technology/workshop on trustworthy elections*, (pp. 12–12). USENIX Association.

Fishbein, M. & Ajzen, I. (1975). *Belief, attitude, intention and behavior: An introduction to theory and research*.

Ganley, D. & Lampe, C. (2006). How deep runs the karma? structural holes and social capital in an online community. *MWAIS 2006 Proceedings*, 25.

Hall, J. (2006). Transparency and access to source code in electronic voting. In *Proceedings of the USENIX/Accurate Electronic Voting Technology Workshop 2006 on Electronic Voting Technology Workshop*, (pp. 8–8). USENIX Association.

Hosp, B. & Vora, P. (2008). An information-theoretic model of voting systems. *Mathematical and Computer Modelling*, *48*(9-10), 1628–1645.

Kim, S.-M. & Hovy, E. (2006). Extracting opinions, opinion holders, and topics expressed in online news media text. In *Proceedings of the Workshop on Sentiment and Subjectivity in Text*, (pp. 1–8). Association for Computational Linguistics.

Kisilevich, S., Rohrdantz, C., & Keim, D. (2010). "beautiful picture of an ugly place". exploring photo collections using opinion and sentiment analysis of user comments. In *Computer Science and Information Technology (IMCSIT), Proceedings of the 2010 International Multiconference on*, (pp. 419–428). IEEE.

Ku, L., Liang, Y., & Chen, H. (2006). Opinion extraction, summarization and tracking in news and blog corpora. In *Proceedings of AAAI-2006 Spring Symposium on Computational Approaches to Analyzing Weblogs*, (pp. 100–107).

Kulkarni, A. V., Aziz, B., Shams, I., & Busse, J. W. (2009). Comparisons of citations in web of science, scopus, and google scholar for articles published in general medical journals. *Jama*, *302*(10), 1092–1096.

Lazarsfeld, P. & Katz, E. (1955). Personal influence. *New York*, 174.

Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis Lectures on Human Language Technologies*, *5*(1), 1–167.

Loncke, M. & Dumortier, J. (2004). Online voting: a legal perspective. *International Review of Law, Computers &# 38; Technology*, *18*(1), 59–79.

Luskin, R. & Fishkin, J. (2005). Deliberative polling, public opinion, and democracy: the case of the National Issues Convention. In *Slightly revised from a paper presented at the annual meeting of the American Political Science Association, Boston, MA, September 2-6, 1998*.

Macdonald, C., Ounis, I., & Soboroff, I. (2007). Overview of the TREC 2007 blog track. In *Proceedings of TREC 2007*.

Merriam-Webster (2004). *Merriam-Webster's collegiate dictionary*. Merriam-Webster.

NASED, N. S. E. D. A. (2002). Voting system standards, volume 1.

Othman, M., Hassan, H., Moawad, R., & El-Korany, A. (2014). Opinion mining and sentimental analysis approaches: A survey. *Life Science Journal*, *11*(4).

Pang, B. & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, *2*(1-2), 1–135.

Phillips, E. (2011). Toward a universal sentiment taxonomy: A multi-language analysis of sentiment traces in a large blog corpus.

Pitt, J., Kamara, L., Sergot, M., & Artikis, A. (2005). Formalization of a voting protocol for virtual organizations. In *Proceedings of the fourth international joint conference on Autonomous agents and multiagent systems*, (pp. 373–380). ACM.

Popoveniuc, S., Kelsey, J., Regenscheid, A., & Vora, P. (2010). Performance requirements for end-to-end verifiable elections. In *Proceedings of the 2010 international conference on Electronic voting technology/workshop on trustworthy elections*, (pp. 1–16). USENIX Association.

Post, R. C. (1990). The constitutional concept of public discourse: outrageous opinion, democratic deliberation, and hustler magazine v. falwell. *Harvard Law Review*, 601–686.

Prevost, A. & Schaffner, B. (2008). Digital divide or just another absentee ballot? Evaluating internet voting in the 2004 Michigan democratic primary. *American Politics Research*, *36*(4), 510.

Reisenzein, R. (2006). Emotions as metarepresentational states of mind. *Cybernetics and Systems*, *2*, 649–653.

Rivest, R. & Smith, W. (2007). Three voting protocols: ThreeBallot, VAV, and Twin. In *Proceedings of the USENIX Workshop on Accurate Electronic Voting Technology*, (pp. 16–16). USENIX Association.

Somasundaran, S. (2010). *Discourse-level relations for Opinion Analysis*. PhD thesis, University of Pittsburgh.

Stark, P. B. (2010). Super-simple simultaneous single-ballot risk-limiting audits. In *Proceedings of the 2010 Electronic Voting Technology Workshop/Workshop on Trustworthy Elections (EVT/WOTE'10). USENIX*.

Stenbro, M. (2010). A survey of modern electronic voting technologies.

Svensson, J. & Leenes, R. (2003). E-voting in Europe: Divergent democratic practice. *Information Polity*, *8*(1, 2), 3–15.

Tang, H., Tan, S., & Cheng, X. (2009). A survey on sentiment detection of reviews. *Expert Systems with Applications*, *36*(7), 10760–10773.

Teague, V., Ramchen, K., & Naish, L. (2008). Coercion-resistant tallying for STV voting. In *Proceedings of the conference on Electronic voting technology*, (pp. 1–14). USENIX Association.

Walsh, D. & Downe, S. (2005). Meta-synthesis method for qualitative research: a literature review. *Journal of advanced nursing*, *50*(2), 204–211.

Wiebe, J. & Deng, L. (2014). An account of opinion implicatures. *CoRR*, *abs/1404.6491*.

Wiebe, J., Wilson, T., & Cardie, C. (2005). Annotating expressions of opinions and emotions in language. *Language resources and evaluation*, *39*(2-3), 165–210.

Wilson, T. A. (2008). *Fine-grained subjectivity and sentiment analysis: recognizing the intensity, polarity, and attitudes of private states*. ProQuest.

Xu, R., Wong, K.-F., & Xia, Y. (2007). Opinmine–opinion analysis system by cuhk for ntcir-6 pilot task. In *Proceedings of the 6th NTCIR Workshop*.

Zhang, L. & Liu, B. (2011). Identifying noun product features that imply opinions. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: short papers-Volume 2*, (pp. 575–580). Association for Computational Linguistics.

Zhang, M. & Ye, X. (2008). A generation model to unify topic relevance and lexicon-based sentiment for opinion retrieval. In *Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval*, (pp. 411–418). ACM.

Zhang, W., Yu, C., & Meng, W. (2007). Opinion retrieval from blogs. In *Proceedings of the sixteenth ACM conference on Conference on information and knowledge management*, (pp. 831–840). ACM.

CHAPTER 4. FOO#THAT AND #THIS+++: A STRUCTURED SENTIMENT USAGE STUDY

In preparation for submission to the proceedings of The Conference on Advances of Social Media Keyword-Hashtag Networks 2015.

Erin Mikel Phillips[1]

Abstract. Hundreds of millions of opinions are being expressed in text each day through social media. However, reliably extracting an opinion author's true intent through opinion mining is difficult. The semantics of opinion are seldom expressed so as to be completely intelligible to an algorithmic classifier. This paper explores the use of a structured sentiment syntax for opinion encoding. 300 students were given opinion encoding and decoding tasks using a hypothetical structured sentiment syntax. Encoding and decoding latency, re-encoding consistency, and general attitudes about structured sentiment were all captured. Structured sentiment yielded improved response times, learning and priming effects and reencoding accuracy at levels well above chance. Comments from subjects included both passionate rejection and effective appropriation of symbolic representations. This exploratory paper provides a basis for much further inquiry into structured forms of opinion expression in text.

[1]Primary researcher and author.

4.1 Introduction

Each day, the numerous and prolific participants in social media encode and publish more than 600M messages for the rest of the world to consider—through just one platform: Twitter. While participation in social media is not demographically uniform (Rainie, 2006; Duggan & Brenner, 2013; Simmons, 2014), in a tangible sense we no longer have to ask: “What do the people have to say?” History has arrived at a place where access to public sentiment does not present a barrier to understanding public sentiment. However, the vagaries of written language have highlighted the fact that understanding public sentiment is a difficult goal to achieve. To date, no one can answer the question: “What do the people mean by what they say?”

It is this second question, about meaning, which motivates opinion mining and sentiment analysis researchers today. With petabytes of social media content to analyze, the hope is that an algorithmic answer to the question of meaning will emerge. Such a solution would enable reliable estimates of what is on the mind and in the hearts of the public.

Opinion mining and sentiment analysis is a recent scholarly discipline, emerging coincident with the adoption and diffusion of read/write Internet applications and platforms.

In 2001 a small number of works examined various methods of extracting sentiment information from free text (Das & Chen, 2001; Pang, Lee & Vaithyanathan, 2002; Turney, 2002; Nasukawa & Yi, 2003; Yi, Nasukawa, Bunescu & Niblack, 2003). At that time, the inventor of the Worldwide Web foresaw the need to shift from the simple tags of HTML to more generalized forms of knowledge representation in what was called the “Semantic Web” (Berners-Lee, Hendler, Lassila & others, 2001). Important work in salience, which is the basis of much of the opinion mining and sentiment analysis research today, can be traced to Cover & Thomas (1991) and others from the immediately-pre-Internet era.

It has been 24 years since Cover & Thomas (1991) and 14 years since Das & Chen (2001). The intensity of academic and commercial interest in opinion mining and sentiment analysis has followed a growth curve similar in shape to that of social media itself. The so-called "land rush" for public sentiment began in 2006, according to Pang & Lee (2008). However, in tribute to the complexities of human attitude and the challenges and limitations of written language, substantial problems remain in reliably extracting basic kinds of sentiment orientation from free text. Liu (2012, p. 13) laments after a lengthy listing of challenges in opinion mining and sentiment analysis, summarized the state of affairs: "These issues all present major challenges. In fact, these are just some of the difficult problems."

4.1.1 Problem

The following examples, from a Twitter corpus used by Bizau et al. (2011), highlight the challenges listed in Liu (2012). In this relatively recent work, a panel of human annotators classified the comments in Examples 4.1 to 4.3 about the movie "Inception." Those annotations became the test data set or "gold standard"—sometimes called "ground truth" used to evaluate the effectiveness of an automated sentiment orientation classifier (Bizau et al., 2011).

Example 4.1. Classified as POSITIVE.

"@shipperfriendly: Haha, love the gif. My brain was trying to understand Inception, I didn't have time or..."

Example 4.2. Classified as NEGATIVE.

"and gahhh still need to see Inception like 3 years ago :("

Example 4.3. Classified as NEGATIVE.

"i got inception on the dvd quality side, but noone to watch it with :("

It appears that in Examples 4.1 to 4.3, the annotators were incorrect. A classifier which reliably replicated the annotated values for Examples 4.1 to 4.3 would also be incorrect. Though being incorrect, the classifier would report high accuracy values. Ku et al. (2006) found that complete agreement among a 3-annotator panel hovered in the 50-60% range on 3-state sentiment orientation for word, sentence, and document-level classification. Ku et al. (2006, p. 2) states, "From the analyses of inter-annotator agreement, we find that the agreement drops fast when the number of annotators increases ... the majority of annotation is taken as the gold standard for evaluation. If the annotations of one instance are all different, this instance is dropped." Using annotators is an established and reliable method of conducting research. However, it appears that the field of opinion mining and sentiment analysis presents a problem space too complex for simplistic approaches to defining ground truth.

The States, Processes, Effects, and Quality (SPEQ) model for opinion mining and sentiment analysis (Phillips, 2015) helps frame the problem (and opportunity) more precisely. Below is a portion of SPEQ which focuses on the encoding and subsequent interpretation of opinion:

- An intended opinion is encoded into a recorded opinion.
- The quality measure for encoding, is usability. Usability is measured by the bias in the encoder and the error introduced into the recorded opinion by the encoder for the given intended opinion.
- The recorded opinion is decoded into an interpreted opinion.
- The quality measure for decoding, is reliability. Reliability is measured by the bias in the decoder and the error introduced into the interpreted opinion by the decoder for the given recorded opinion.

SPEQ prescribes two questions for opinion mining and sentiment analysis researchers, such as Bizau et al. (2011):

1. What is the bias and error in the encoding?

2. What is the bias and error in the decoding?

In (2) above, is the essence of opinion mining and sentiment analysis. However, what is the "truth" of the statement in Examples 4.1 to 4.3? How "reliable" is the decoder used by Bizau et al. (2011)—or any other free-text decoder, really, if the "gold standard" is assumed to have zero bias and zero error, but doesn't?

The purpose of opinion mining is to mine or extract opinion from free text. However, opinion itself is a "private state" (Post, 1990). Therefore, opinion mining and sentiment analysis research is actually trying to extract "intended opinions" from "recorded opinions" to create "interpreted opinions." It is this chain-of-custody problem which presents a substantial problem for miners of free text. It also presents a tremendous opportunity for the use of a structured opinion encoding syntax.

The following is a replay of the above example using the language of SPEQ:

- Subject encodes an intended opinion into a recorded opinion.

- Annotators decode the subject's recorded opinion into an interpreted opinion. which becomes ground truth.

- Opinion mining algorithm decodes the subject's recorded opinion into an interpreted opinion.

- Opinion mining algorithm is accurate 85% of the time, when compared to ground truth.

- Note: inter-annotator agreement rate was 70%.

The above analysis assumes that we can identify the specific target of the opinion statement—a tenuous assumption in the general case. The use of SPEQ as a framework clarifies some of the assumptions being made about how much is known about what is being measured. The use of a structured sentiment encoding syntax may eliminate the decoding difficulties entirely.

The Universal Voting Markup Language (UVML) (Phillips, 2011b) is a structured sentiment encoding syntax. A syntax specification for UVML is included in Appendix C. Consider the following example where UVML is used. In Example 4.4, the text in Example 4.3 has been altered to include a simple UVML annotation.

Example 4.4. Classified POSITIVE comment about the movie "Inception."

"i got #inception+++ on the dvd quality side, but noone to watch it with :("

Herein lies the potential for structured sentiment expression. What is the difference between Example 4.4 and Example 4.3? Is this a positive or negative statement about the movie "Inception"?

SPEQ is also very helpful in understanding why the opinion mining process is different in Example 4.4. To be precise, with structured sentiment, there is no opinion mining—only "opinion reading", or in a the parlance of SPEQ, "voot counting."

Theoretically, a well-designed structured representation of sentiment has bias and error terms which tend toward zero for both encoding end decoding. In that case, "intended opinion" equals "recorded opinion" which equals "interpreted opinion."

4.1.2 Purpose

A serviceable estimate is that 50% of the social media content contains statements of opinion (Macdonald et al., 2007; Phillips, 2011a). Given the immaturity of context-less opinion mining algorithms, it is safe to say that a large proportion of the 300M opinions expressed on Twitter today, are not counted by any social voting, or vooting, system. Structured opinion representations could improve access to those uncounted opinions. However, little work has been done to formalize the expression of opinion in free text.

UVML is an encoding scheme designed for this purpose. The purpose of this paper is to examine differences in the human experience between expressing opinion and consuming opinion expressions in both the unstructured form shown in Examples 4.1 to 4.3 and the structured form demonstrated in Example 4.4. The following research questions guide this inquiry.

Research Question 4.1. *For users of social media, what are the similarities and differences between expressing and consuming an opinion in words versus a structured sentiment syntax?*

Research Question 4.2. *How, if at all, do the effects identified in Research Question 4.1 differ across subject demographic or experiential categories?*

Research Question 4.3. *How, if at all, do the effects identified in Research Question 4.1 differ across contextual factors related to the individual opinions expressed, such as the type of topic or order of presentation?*

4.2 Background

Structured sentiment notations are not used today in social media. However, a few attempts have been made both historically, and in the social-media age, to add structure to free-form text. The differentiator amongst these schemes is the extent to which they reflect the semantics of opinion as defined by Baker et al. (1998).

4.2.1 Markup Languages

Markup is meta-data, which clarifies the intentions of the document author. A markup language is a specification for a kind of markup. There are three types of textual markup: presentation (how to render), procedural (how to process), and descriptive (what type). The concept of a markup language was not new when Tim Berners-Lee received funding at CERN to develop HTML in 1990. Goldfarb & Rubinsky (1990) describes how he and his team from IBM developed a notation for standardized textual markup throughout beginning in the 1960s. Markup sequences such as `<TITLE>The Title..</TITLE>` had been in use by engineers, academics, and publishers for nearly two decades prior to 1990. The difference between the conventional notions of markup and a structured sentiment syntax are numerous. However, the primary difference is target audience.

In the case of conventional markup, the intended audience is a text processing system, which will use the embedded tags to perform actions on the text according the tags. In the case of a sentiment syntax, the intended audience includes a text processing or vooting system. However, the audience is also human.

Example 4.5. An example opinion using a hypothetical semantic markup language.

"i got <OPINION value="like"><TARGET type="movie">inception</TARGET></OPINION> on the dvd quality side, but noone to watch it with <FEELING value="sad">:(</FEELING>"

Markup languages have the potential to fully express the semantics of opinion as defined by Baker et al. (1998). However, few social media users would take the time to type in the text of Example 4.5. The audience is clearly intended to be a text processing system.

4.2.2 Emoticons/Emoticons

Emojis, or the cute little syntactic faces which dot the social media landscape, have been studied (Read, 2005; Go, Huang & Bhayani, 2009; Aoki & Uchida, 2011). In Go et al. (2009), the following mapping was provided:

- POSITIVE emoticons :) , :-) , :) , :D , =)
- NEGATIVE emoticons :(, :-(, : (

The limitation with emoticons is that they lack the essential features of opinion semantics. Consider the use of emoticons in Example 4.3. The **:(** is a representation of how the author feels about the proposition that he or she does not have anyone with which to watch the movie. Emojis and emoticons are not anchored to either a particular target or a defined opinion.

4.2.3 The Universal Voting Markup Language (UVML)

UVML (Phillips, 2011b), was an effort to codify a set of hashtag annotations to encode opinion type, magnitude, and direction information in free text. Phillips (2011b) identified five (5) types of opinions: quality, importance, outlook, support or opposition, and likelihood. Each of these types of opinions was assigned a symbol for encoding. For example, **`#coke*****`** is a statement about the quality of Coke, being "among the best." On the other hand, **`#obama+++`**, is a statement of support and **`#obama---`** is a statement of opposition.

4.2.4 Oofoo

An unpublished extension to UVML included an OO and FOO prefix for hashtags, to represent a visceral reaction to the opinion target. For example, **`oo#beets`**, is a positive affective response to beets. On the other hand, **`foo#broccoli`**, is a negative affective response to broccoli. This allows for emphasis, what Baker et al. (1998) calls "manner." The specification allows for repeating Os. So, ooooo#beets, is a stronger representation of a positive affective response than **`oo#beets`** as **`foooooo#broccoli`** is more strongly negative.

4.3 Methods

Data collection for this study was done using a customized application designed explicitly for the purpose. The application was entitled "Survey on Sentiment in Social Media", and subsequently designated "S3." Where helpful within in this paper, the moniker "S3" is used as a reference to this application.

S3 presented subjects with two types of tasks to complete. The first type was to respond to survey questions. The second, perform a series of subtasks involving encoding and decoding opinions. Through the course of the subject interactions with S3, many events and variable values were captured. Because no similar examples of investigations into structured sentiment were found in a review of the literature, the null hypotheses listed in the results are not derived from well-defined theoretical constructs. Instead, the methodology followed in this paper is on more of exploration. This paper presents a "first look" into end-user reactions to and performance using structured sentiment. Interesting features in this data are examined through a null hypothesis. However, some are presented simply as findings.

4.3.1 Participants

Approximately 495 undergraduate students in Advertising (ADVRT 230) and Media and Communications (JLMC 101) courses were given an opportunity to participate in this study. The following email text was sent to students in these two classes.

> Thank you for your willingness to participate in the "Survey on Sentiment in Social Media." The purpose of this study is to gain additional insights into two questions: 1) Is sentiment expression an important part of social media usage? 2) What forms of sentiment expression in social media work the best? The survey consists of 44 questions, and should take less than 20 minutes to complete.

Table 4.1: Participants and S3 application session summary

	JLMC 101	ADVRT 230	Unknown	TOTAL
POPULATION	282	213	-	**495**
PARTICIPANTS	267	135	41	**443**
Participation Rate				*89.5%*
EXCLUSIONS				
.. (1) Data corruption	0	0	-1	**0**
.. (2) Replaced by retry	0	-1	-3	**-4**
.. (3) Patently inattentive	-13	-7	-2	**-22**
.. (4) Attempted bookmark use	-50	-15	-11	**-76**
.. (5) Testing	0	0	-3	
.. (6) Inattentive (>5 minutes)	-8	-5	0	**-11**
.. (7) Completely empty	-4	-5	-2	**-11**
USABLE RESPONSES	**191**	**102**	**20**	**313**
Useable Response Rate				*63.2%*

Each student who participated received nominal credit (0.5%) toward their grade for participating. A summary of the subject responses is provided in Table 4.1. Subjects who participated in the study were instructed to only participate once if enrolled in both classes. There were 34 of these dually enrolled subjects, and their results are included in the JLMC 101 scores, for simplicity.

The overall participation rate was high, according to the instructors—perhaps an indication of a type of populist energy around things connected to social media. Of the 495 possible subjects, 443 chose to participate: a participation rate of 89.5%.

Exclusions. Some responses were excluded from subsequent analysis for a variety of reasons, shown in Table 4.1. Some exclusions merit explanation. Exclusion #3 above, patently inattentive responses—of which there were 22, were survey responses where subjects attested to usage of a hypothetical social media platform, "SocialMe." Survey responses in which the subject indicated that they actively use the hypothetical service were excluded. Presumably those subjects were not paying close attention to the question.

A large number of subjects attempted to bookmark timed elements of the experiment, presumably to return at a later time. However, bookmarking created many incomplete responses or excessive response latency values. There were 76 responses excluded because of bookmarking—exclusion #4, above. Exclusion #5, inattentive, by response latency >5 minutes for a single question, removed 11 responses from consideration. Presumably, these individuals had to step away from the S3 application in the middle of their session. Lastly, there were 11 empty surveys— with no supplied responses to any questions or tasks. These empty responses were also excluded from the analysis by exclusion rule #7—presumably these subjects only wanted the credit, which was fine as all questions were optional.

Useful Responses, Age, and Gender. Of the original 443 participants, 130 responses were excluded from further analysis from exclusion rules #1-7 shown in Table 4.1, leaving 313 usable responses for analysis—63.2% of the total available population. Of the 313 useful responses, 251 supplied a value for age. The range of ages was 18 to 31, with the mean age of 20.3 years, with a standard deviation of 1.6 years. Of the 252 who supplied a value for gender, 178 identified themselves as female, 74 as male.

4.3.2 Materials

S3, the application designed for this study, was the subject user interface and the source of all data collected. The application was written using a proprietary programming language developed by the primary investigator and hosted in Heroku's virtualization environment. A detailed description of the application is included in Procedures, Section 4.3.3. In summary, through S3, the study asked subjects to complete nine (9) distinct tasks, organized into four (4) parts. The presentation of tasks in S3 corresponded to the parts and tasks in the study design. The parts and tasks are listed below and discussed in more detail with screenshots in Section 4.3.3:

- Part 1 - About
  - Tasks 1-3. About: reviews some basic terminology relevant to the study.
- Part 2 - Consent
  - Task 4. Consent: discloses risks and terms, and captures subject agreement.
- Part 3 - Survey and Encoding/Decoding Tasks
  - Task 5. Provide age, gender, and social media usage experience.
  - Task 6. Encode their opinions on various topics by choosing from selection of options, presented either as words or opinions pre-encoded using a structured sentiment syntax such as OOFOO, PLUS/MINUS, or STARS.
  - Task 7. Decode someone else's opinion which has been encoded using a structured sentiment syntax.
  - Task 8. Record their previous exposure to the use of structured sentiment in social media.
  - Task 9. Express an opinion using whatever form of opinion representation he or she chooses.
- Part 4 - Acknowledgement and Thank you.

As each subject performs tasks 5-9, a number of variable values are captured or calculated. Though many variable values were captured by S3, only those used in the paper will be discussed, unless the omission requires explanation. For coherence, the variables relevant to subsequent analysis are discussed briefly adjacent to the task description. Variables are shown in italics, $topicCategory$, for example. Categorical variable values are shown in UPPER-CASE, and each includes a value of UNKNOWN to represent a non-response. The next section, Section 4.4, reviews the findings using the variable values captured by S3.

4.3.3 Procedures

Each potential subject received an email with a URL-link which would take them to the S3 application to participate in the study. Upon clicking on the link, the following application pages were presented to enable the subjects to perform tasks 1-9 listed above.

Task 1 - About Key Terms. Task 1 asked the subjects to review definitions for “sentiment” and “social media” and acknowledge a level of comfort with the terms. The Task 1 page is shown in Figure 4.1 in the context of the entire S3 page template. Subsequent screen shots will only include the non-boilerplate content. Subjects were also advised that they have the opportunity to exit the study application at any time.

Task 2 - About the Study Goals. Task 2, shown in Figure 4.2, presents the objectives of the study in simple terms and asks the subject to review the stated goals, and acknowledge a level of comfort with those goals.

Task 3 - About the Study. Task 3, explains to the subject that the study will consist of approximately 44 questions and should take less than 20 minutes to

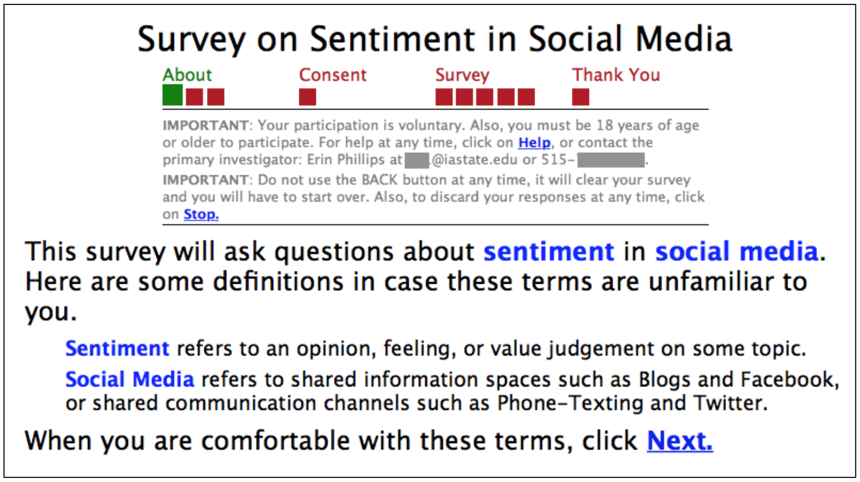


Figure 4.1: Task 1: review relevant terminology

Figure 4.2: Task 2: review the goals of the study

complete—and is shown in Figure 4.3. The subject was asked to make a commitment to the time required, and express an intention to complete the survey.

Task 4 - Consent. The consent task, Task 4, involves electronically signing a consent form consistent with standards for responsible research using human subjects. Signing the consent form was required to continue with the survey. The explanatory text for Task 4 is shown in Figure 4.4. After clicking on the consent link, the subject was shown the consent form and given an opportunity to give

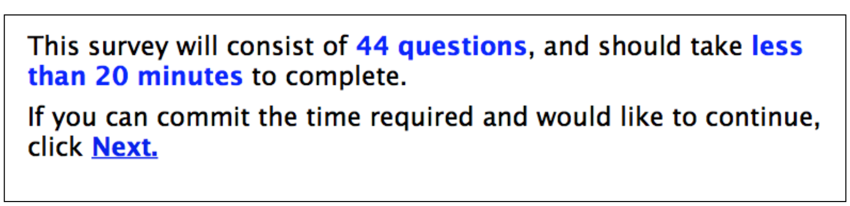


Figure 4.3: Task 3: acknowledge study scope and time required

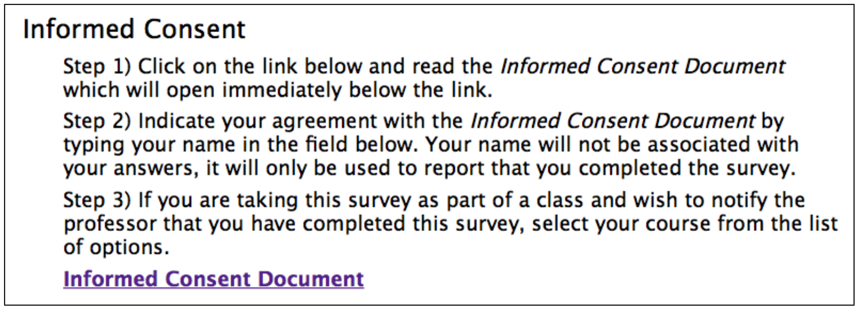


Figure 4.4: Task 4: consent to participate

their name and indicate if they were taking the survey as part of one of the courses participating.

Task 5 - Subject Demographics and Social Media Usage. In Task 5, the subject is asked to supply basic information such as age, gender, and social media usage experience. The S3 Task 5 panel is shown in Figure 4.5. Social media usage experience took the form of asking about the subject's frequency of use of popular social media platforms such as Twitter, Facebook, or Google+. Also included in Task 5 was a question about what percentage of their use of social media include reading the opinions of others, and what percentage included writing opinions.

Task 5 captures a number of important demographic and social media usage variables, including gender, and social media usage by channel. Also, the percentage of opinion content read and written is requested from the subjects. Age was captured in this task, but not used as an independent variable in any analysis because the standard deviation of age across the study population was very small (<1.6). Table 4.2 lists the variables captured in Task 5.

Task 6 - Encode Opinions.

The encoding task, Task 6, is the first instrumented task for subjects. Because learning effects are important phenomena to this study, subjects are not given any

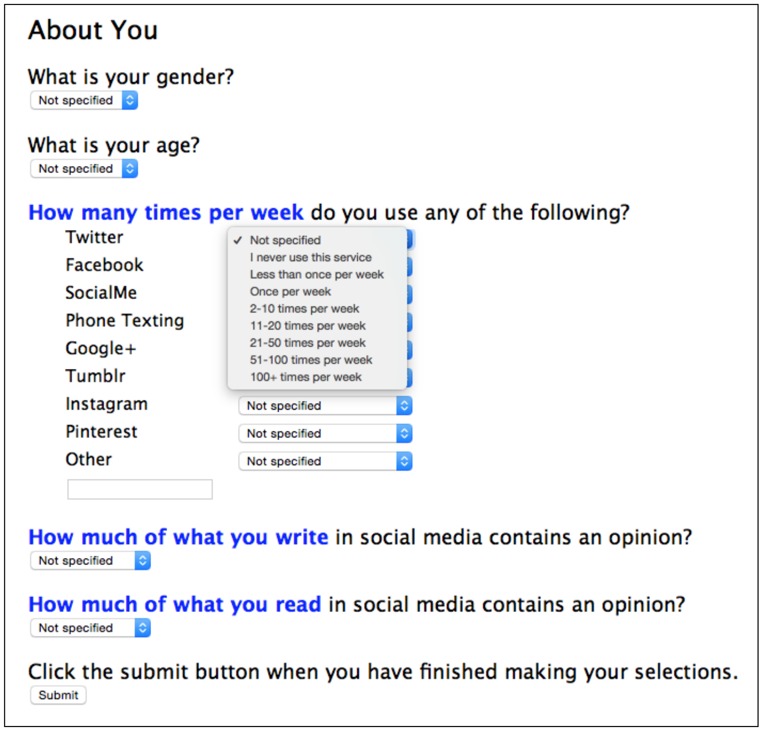


Figure 4.5: Task 5: provide demographics and social media usage

Table 4.2: Task 5 variables captured or calculated by S3

gender. categorical; values UNKNOWN, MALE, FEMALE.
usagePerWeek. descrete ratio; equal to the sum of the lower bound categorical values for each social media service indicated.
opinionReadPercent. categorical; values <11, 11-50, 51-90, and 90+, representing the percent of social media content read by the subject which includes opinions.
opinionWritePercent. same as $opinionReadPercent$, except value is percent of social media content written containing opinion.
opinionReadCount. descrete ratio; equal to $usagePerWeek$ $*$ $opinionReadPercent$ and represents an estimate of the number of opinions read by the subject in the course of a week.
opinionUserClass. categorical; values of READER, BALANCED, or WRITER. This value is calculated for each subject as follows: if $opinionReadPercent$ $>$ $opinionReadPercent$ then READER; if $opinionReadPercent$ $<$ $opinionWritePercent$ then WRITER; else BALANCED.
socialMediaUsageLevel. categorical; values of LIGHT, MODERATE, or HEAVY. This value is calculated for each subject as follows: if percentile($usagePerWeek$) is <25 then LIGHT; <75 then MODERATE; else HEAVY.

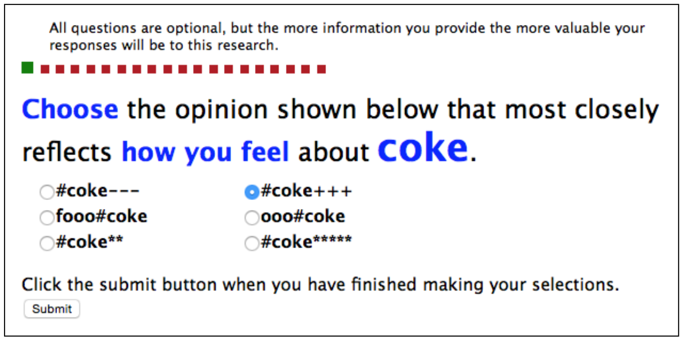


Figure 4.6: Task 6: encode opinion using a symbolic representation

instructions, beyond the question itself: "Choose the opinion shown below that most closely reflects how you feel about {target}." {Target} is one of the ten possible topic (*topic*) values, spanning four categories (*topicCategory*) described in Table 4.3. The options for choosing an opinion are presented in the form of words or presented using three different structured opinion encoding schemes (*scheme*): OO/FOO, PLUS/MINUS, and STARS. Two polar opposite values are given for each.

Each of the ten topics is presented twice to the subject, once with WORD options as in Figure 4.7, and once with SYMBOLIC options as shown in Figure 4.6. The order of topics presented in Task 6 is randomized. The order of which comes first, WORD or SYMBOLIC, is also randomized.

In the example screen shots, the subject was asked about his or her opinion about COKE. In the first presentation of COKE—item #1 of 20, shown in Figure 4.6, the subject was presented with options using a symbolic opinion syntax. The small squares across the top of the panel represent where in the sequence of subtasks the subject is. In the second presentation of the target COKE—item #15 of 20, shown in Figure 4.7, the subject was given choices using words.

In addition to capturing the subject's selections, the elapsed time is captured. For each encoding attempt, the elapsed time is the number of milliseconds from the time the panel is fully rendered to the time the subject clicks on SUBMIT—as recorded on the client web browser.

Table 4.3: Task 6 variables captured or calculated by S3

encodeTime. descrete ratio; the span of time in milliseconds between the moment the page is fully rendered and the moment that the subjects clicks submit.
encodeAttempts. descrete ratio; the number of attempts made by the subject to encode an opinion, signified by clicking submit.
encodeType. categorical; values of WORD or SYMBOL determined by which type of opinion representations are gfiven to the subject to choose from when encoding.
encodeRate. continuous ratio; equal to $(\frac{encodeAttempts}{\sum encodeTime}) * 60,000$
encodeRateWords. continuous ratio; the $encodeRate$ when $encodeType$ is WORD.
encodeRateSymbols. continuous ratio; the $encodeRate$ when $encodeType$ is SYMBOL.
topicCategory. categorical; values of CONCEPT, ISSUE, PEOPLE, and PRODUCT.
topic. categorical; values of [CONCEPT] FAMILY, HONESTY, LIFE; [ISSUE] ABORTION, DEATHPENALTY, GUNRIGHTS; [PEOPLE] OBAMA, REPUBLICANS; [PRODUCT] COKE, PEPSI.
scheme. categorical; values of OO/FOO, PLUS/MINUS, and STARS.
opinionWords. categorical; values of SUPPORT/OPPOSE, LIKE/DISLIKE, IMPORTANT/UNIMPORTANT, LIKELY/UNLIKELY, OPTIMISTIC/PESSIMISTIC.
initialPresentation. categorical; values of YES and NO, reflects whether or not this particular encoding response was the inital presentaiton of the $topic$.
previousAttempts. descrete ratio; the number of previous attempts for the $encodeType$ (WORD or SYMBOL) associated with the encoding task.
experientialPhase. categorical; LEARNING and APPLYING, for values of $previousAttempts$ which are 0-2 and 3+ respectively.

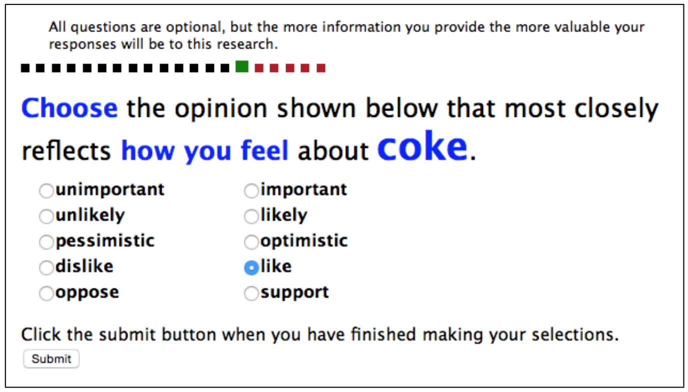


Figure 4.7: Task 6: encode opinion using words

Task 7 - Decode Opinions.

After the 20 encoding subtasks, the subject is asked to do some decoding of opinions of others encoded in symbolic notations. The user is presented with 15 decoding subtasks, one opinion of another person in each. Of the 15 decoding subtasks, 10 are random selections. The remaining five decoding subtasks are re-presentations of the subjects encoded opinions from Task 6. For example, in Figure 4.6 and Figure 4.7, the subject is presented with a request for their opinion about COKE: #coke+++ and "like", respectively. In Task 7, the subject would at some point be asked to decode the opinion of someone else—except the subject is presented his or her response from Task 6. The panel shown in Figure 4.8 shows how this appears to the subject for the target COKE. A number of variables and derived values are captured by S3 as the subject performs Task 7, and these are described brief in Table 4.4.

In addition to capturing the subject's selections, the elapsed time is captured. For each decoding attempt, the elapsed time is the number of milliseconds from the time the panel is fully rendered to the time the subject clicks on the "submit" button—as recorded on the client web browser.

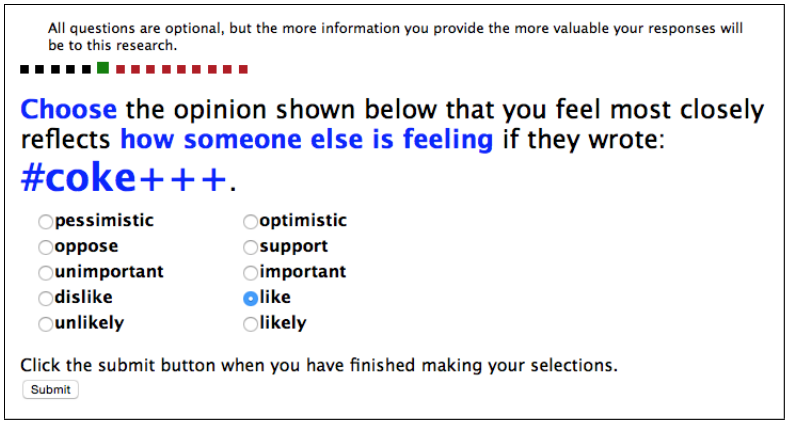


Figure 4.8: Task 7: decode opinion encoded with symbols

Table 4.4: Task 7 variables captured or calculated by S3

decodeTime. descrete ratio; the span of time in milliseconds between the moment the page is fully rendered and the moment that the subjects clicks submit.
decodeAttempts. descrete ratio; the number of attempts made by the subject to decode an opinion, signified by clicking submit.
decodeRate. continuous ratio; equal to $(\frac{decodeAttempts}{\sum decodeTime}) * 60,000$
decodeMatchToEncode. categorical; MATCH or MISSMATCH, depending on whether or not the Task 7 value matches the Task 6 value for the same target

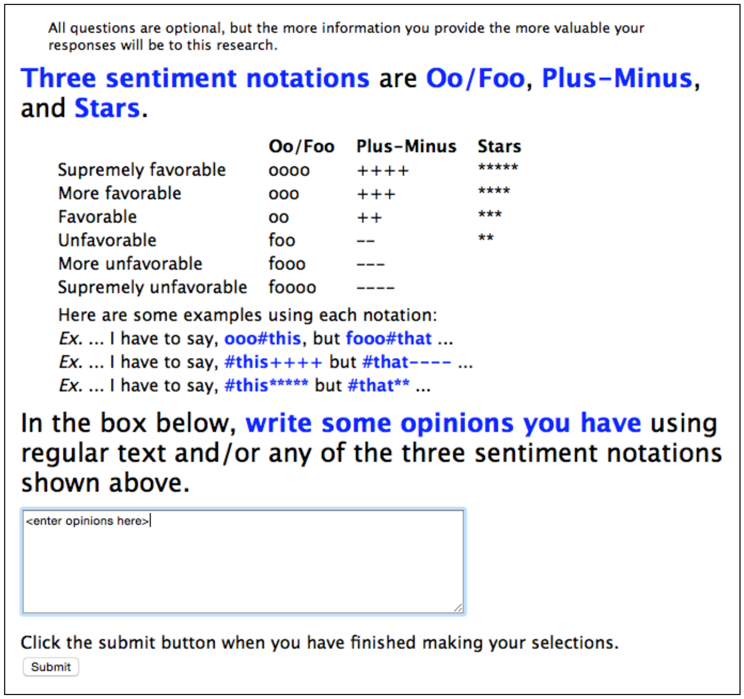
Figure 4.9: Task 8: provide previous experience with and prospective use of structured sentiment

Table 4.5: Task 8 variables captured or calculated by S3

exposureFrequency. categorical; values of NEVER, RARELY, SOMETIMES, FREQUENTY.
likelihoodDecoding. categorical; values of EX-UNLIKELY, UNLIKELY, NEITHER, LIKELY, EX-LIKELY.
likelihoodEncoding. categorical; values of EX-UNLIKELY, UNLIKELY, NEITHER, LIKELY, EX-LIKELY.

Task 8 - Previous Experience with Structured Sentiment.

After completing Task 7, subjects were presented with three (3) questions shown in Figure 4.9. Subjects were asked what kind of exposure they previously had to symbolic opinion encoding schemes. Subjects were also asked to think prospectively. Specifically, subjects were asked how likely he or she is to correctly decode opinions written using an opinion encoding scheme; and, how likely to use an encoding scheme when writing an opinion in the future. The variable values captured by S3 in Task 8 are shown in Table 4.5.



All questions are optional, but the more information you provide the more valuable your responses will be to this research.

Three sentiment notations are **Oo/Foo**, **Plus-Minus**, and **Stars**.

	Oo/Foo	Plus-Minus	Stars
Supremely favorable	oooo	++++	*****
More favorable	ooo	+++	****
Favorable	oo	++	***
Unfavorable	foo	--	**
More unfavorable	fooo	---	
Supremely unfavorable	foooo	----	

Here are some examples using each notation:
Ex. ... I have to say, **ooo#this**, but **fooo#that** ...
Ex. ... I have to say, **#this++++** but **#that----** ...
Ex. ... I have to say, **#this***** but **#that**** ...

In the box below, **write some opinions you have** using regular text and/or any of the three sentiment notations shown above.

<enter opinions here>

Click the submit button when you have finished making your selections.

Submit

Figure 4.10: Task 9: enter an opinion using text or an opinion encoding syntax

Task 9 - Comments on Anything.

The final task for subjects was to use a free-form comment box to enter a comment on anything about which they have an opinion. As shown in Figure 4.10, above the comment box was a detailed explanation of the sentiment encoding symbol schemes they had encountered during their participation in the study. Entering a comment was optional.

After clicking SUBMIT on the panel for Task 9, a closing acknowledgment and thank you were provided to the subject to indicate completion of the survey. The next section lists the variables gathered from the above procedures, in preparation for the statistical analysis of the questions posed by Research Questions 4.1 to 4.3. The variable values captured by S3 in Task 9 are shown in Table 4.6.

4.3.4 Disclosures

Table 4.6: Task 9 variables captured or calculated by S3

subjectComment. text; a free-form comment or opinion by the subject on anything.
subjectCommentType. categorical; values of NO-COMMENT, OBSERVATION, OPINION-SYM, and OPINION-TEXT.
subjectCommentBias. categorical; NO-COMMENT, NEGATIVE, NEUTRAL, and POSITIVE.

This research was otherwise conducted in accordance with the guidelines published by Iowa State University Institutional Review Board regarding the protection of human participants in the Investigator Handbook. The images of the screens in this paper are exact representations of the what the subjects saw in the course of their participation. The descriptions of the activities in which the subjects engaged as part of this study are described in accordance with how they occurred.

4.4 Results

The following results were obtained in a post-hoc analysis of the data collected by S3. The experiment itself was motivated by Research Questions 4.1 to 4.3, but the explanations in each section below were defined more in terms of the data collected, rather than a rich theoretical construction. This approach is consistent with other works which introduce new domains of inquiry, such as Pang et al. (2002). Pang et al. (2002) did not include a formal hypothesis, but rather explained the methodology for determining the sentiment lexicon for the classifier which was tested:

> "One might also suspect that there are certain words people tend to use to express strong sentiments, so that it might suffice to simply produce a list of such words by introspection and rely on them alone to classify the texts . . . To test this latter *hypothesis*, we asked two graduate students

> in computer science to (independently) choose good indicator words for positive and negative sentiments in movie reviews." (p. 2)

The results presented in this section follow the more informal approach taken by Pang et al. (2002) and others. Each view of the data captured by S3 is prefaced with an explanation of the variables used to construct the view. These variables are discussed the first time they are used. The use of the term "subjects" within this portion of the paper refers to those participants whose data remained after the cleansing process described in Section 4.3.1.

4.4.1 Subject Opinion Experience

The self-reported values for opinion generation and opinion consumption help to describe the "flow" of opinion in the social media eco-system. The variable $opinionUserClass$ was used to characterize the directional of opinion flow.

As shown in Figure 4.11, there is an 8:1 ratio of opinion WRITERS to READERS. This finding suggests that social media is "write-heavy." Moreover, this finding may be important for issues relating to structured sentiment. It emphasizes the importance of having reliable voot decoding services. The services for consumption of opinion are dwarfed by those available for expressing opinion. This imbalance is visible in the large disparity between opinion expression and opinion consumption shown in Figure 4.11 and Figure 4.12 farther below.

A more detailed look at the per subject values for read and write activity is shown in Figure 4.12. Here, the relationship between $opinionReadCount$ and $opinionWriteCount$ is shown grouped by $opinionUserClass$—and the domination of opinion expression over opinion consumption is apparent.

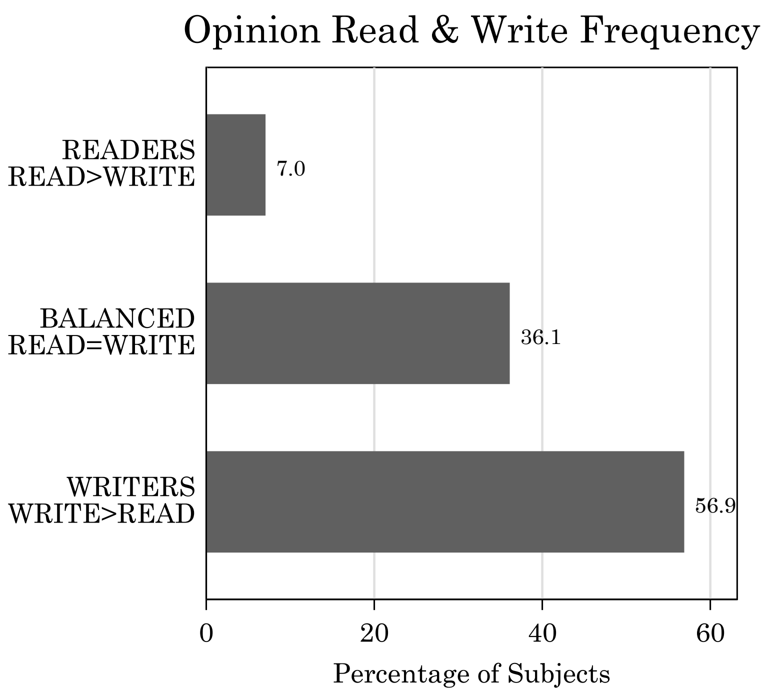


Figure 4.11: Subject counts by *opinionUserClass*

4.4.2 Encoding Rates

One of the most important potential differences between encoding an opinion with words versus encoding an opinion using a symbolic encoding scheme would be the cognitive load required. In pursuing Research Question 4.1 with regard to encoding rates, a null hypothesis can be defined. Response times have been effectively used as representation of cognitive load (Hancock, Thom-Santelli & Ritchie, 2004; Tsur & Rappoport, 2012).

It would be reasonable to assume that first-time users would struggle to use an opinion encoding scheme for which there are only shadowy and incomplete analogs being used today in social media. Therefore, the *encodeRateWords* would be expected to be higher than the *encodeRateSymbol*, i.e., subjects can complete more encoding tasks per minute using words than using symbolic encoding schemes.

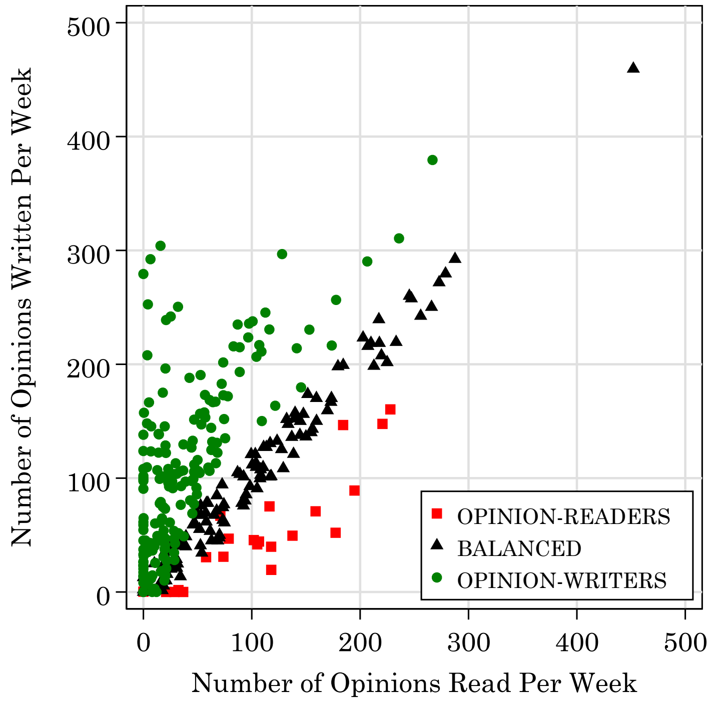


Figure 4.12: Subject $opinionReadCount$ and $opinionWriteCount$ by *opinionUserClass*

This outcome would occur if the cognitive effort required to encode sentiment using a symbolic representation is substantially greater than what it takes to encode an opinion using words. Therefore, it seems reasonable to define a null hypothesis:

Null Hypothesis 4.1. *A subject's encoding rate per minute using words ($encodeRateWords$) will be less than the subject's encoding rate per minute using symbols ($encodeRateSymbols$.)*

A graphic representation of the result from testing Null Hypothesis 4.1 is shown in Figure 4.13. The t-test value for Null Hypothesis 4.1 was $p < 1.0$ $(df = 624)$. Null Hypothesis 4.1 cannot be rejected. Moreover, when Null Hypothesis 4.1 is partitioned using $gender$, the findings were similar for both male and female subjects.

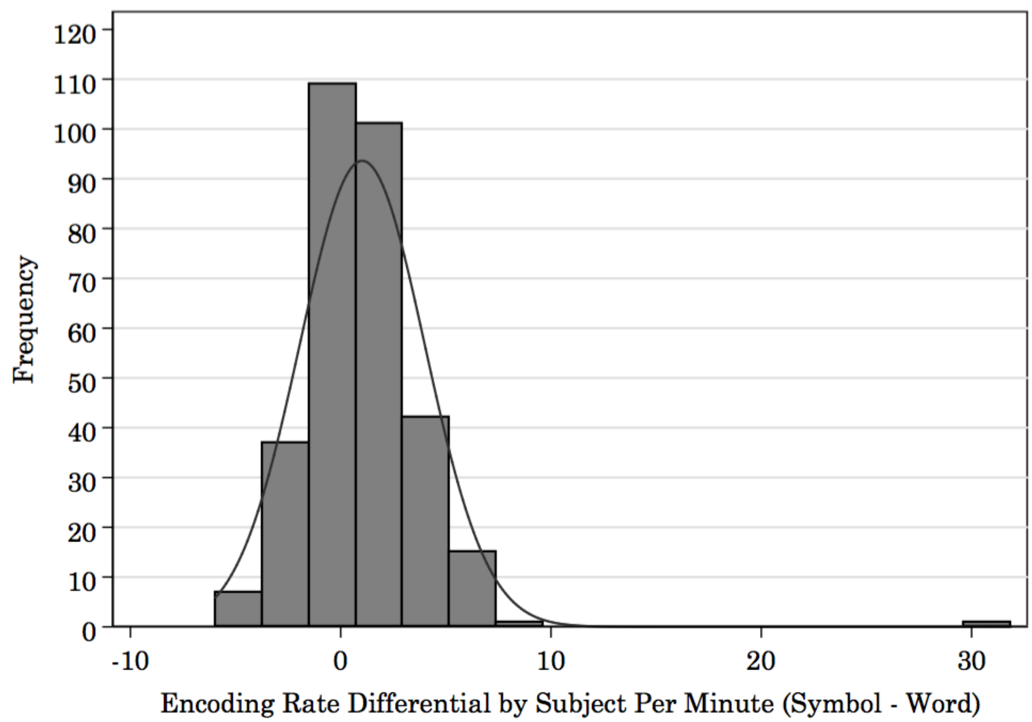


Figure 4.13: Distribution of $encodeRateSymbols$ - $encodeRateWords$

By Gender by Encoding Type.

While the diversity of ages in the population is not sufficient to meaningfully partition the encoding rates by age, gender effects seem plausible—though it is not obvious that the effects would be in any particular direction. Emotional intelligence, reading skills, and language arts scores in women may be higher than those of men. On the other hand, men may be more analytical (on average) than women. Symbolic encoding of private states seems to span both domains. For lack of directly related prior scholarship to push the inquiry in one direction or another, the following null hypotheses are proposed:

Null Hypothesis 4.2. *Female opinion encoding rate per minute using symbols ($encodeRateSymbols$) will be greater than the opinion encoding rate per minute for males using symbols ($encodeRateSymbols$.)*

If it is assumed that language skills dominate the task—the following null hypothesis can be formed:

Null Hypothesis 4.3. *Male encoding rate per minute using words ($encodeRateWords$) will be greater than the encoding rate per minute for females using words ($encodeRateWords$.)*

Using a t-test to compare the mean $encodeRateSymbols$ of the two groups, female (178) and male (74), Null Hypothesis 4.2 can be rejected with $p < 0.03\,(df = 250)$—supporting the contention that males may be able to encode opinions using symbols at slightly higher rates than females. For Null Hypothesis 4.3, using a t-test to compare the mean encoding time using $encodeRateWords$ of the two groups, the null hypothesis cannot be rejected as $p < 0.88\ (df = 250)$ —indicating that there are no meaningful gender effects in the Task 6 encoding of opinions using words in S3. A graph showing these results is shown in Figure 4.14

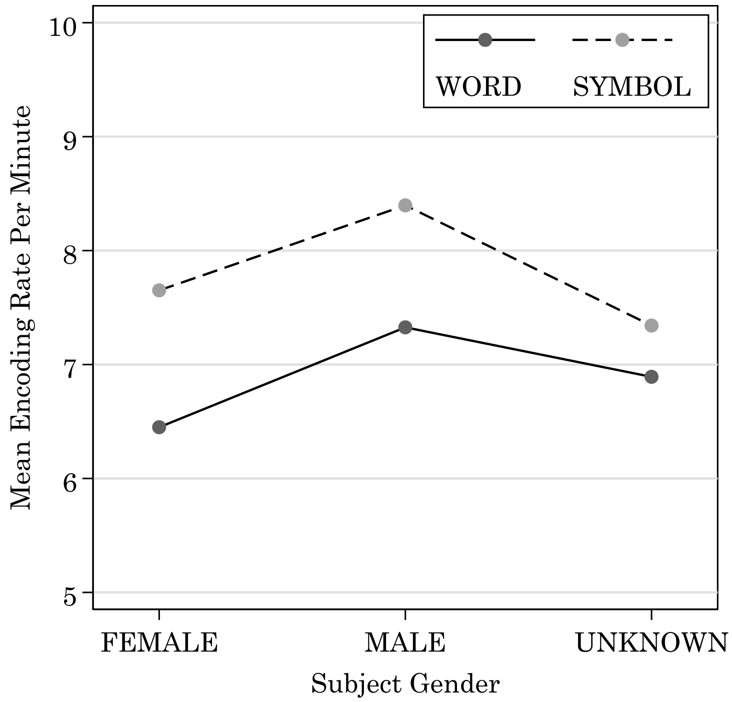


Figure 4.14: Encoding rates by *gender*

By Social Media Usage Levels.

It seems plausible that prolific users of social media will be more adept at using a symbolic opinion encoding scheme than casual users or non-users. Because aggregate social media usage ($usagePerWeek$) is a granular measure, it makes sense to look at the relationship between social media usage and encoding rate per minute using words ($encodeRateWords$) and symbols ($encodeRateSymbols$) through regression.

The following two null hypotheses seem to present themselves given the operational character of the variables. First, it could be posited that there should be little if any encoding rate effects when words are used as social media is not a mediating factor in exposure to words. So, the following null hypothesis seems to be defensible:

Null Hypothesis 4.4. *Social media usage per week ($usagePerWeek$) is a reliable predictor of opinion encoding rate per minute using words ($encodeRateWords$).*

Likewise, it seems plausible that social media usage experience is a useful predictor of subject encoding rate per minute using symbols. Many subjects, no doubt, encode their thoughts using a few hundred emoticons per week. Given this proposition, the following null hypothesis seems to make some sense:

Null Hypothesis 4.5. *Social media usage per week ($usagePerWeek$) is not a reliable predictor of opinion encoding rate per minute using symbols ($encodeRateSymbols$).*

Figure 4.15 shows the results relating to Null Hypotheses 4.4 and 4.5, including a regression line. As shown, social media usage levels are not a good predictor of opinion encoding rates for either words or symbols. We can reject Null Hypothesis 4.4 because $p < 0.41\ (R^2 = 0.0)$. We cannot reject Null Hypothesis 4.5 because $p < 0.31\ (R^2 = 0.0)$.

4.4.3 Encoding Priming Effects

As discussed in Section 4.3.3 under Task 6, each topic was presented to subjects twice: once where the choices were given in words and once when the choices were given using a symbolic opinion encoding scheme. The order of which came first and the number of intervening questions was randomized. Theoretically, then, the fraction of improvement (if any) in the response times between the first presentation and the second presentation for the same target may be attributable to a priming effect—that is, the subject has already determined how he or she feels about the target.

This analysis is important because it gets to the heart of the question regarding the veracity of subject responses to symbolic encoding. If subjects are taking the time to connect the symbolic representations of opinion with the target, then there

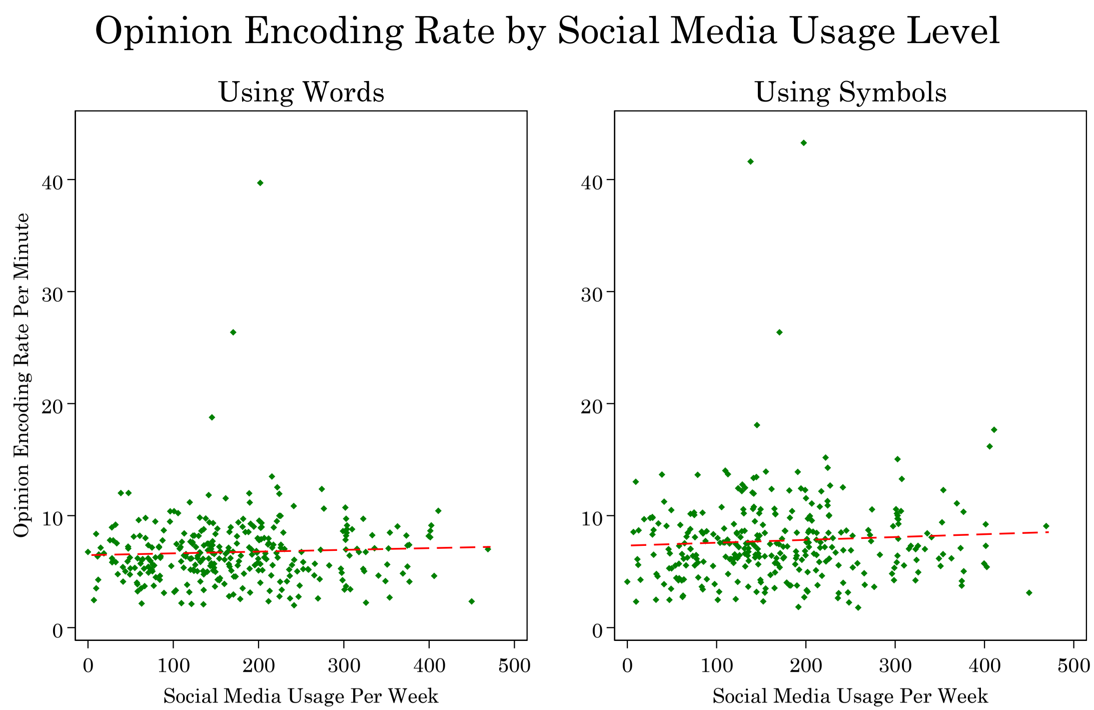


Figure 4.15: Subject *encodeRateWords* and *encodeRateSymbols* by ***usagePerWeek***

should be some priming effects when symbols are given as options first—just as would be expected for words. Using this theoretical construction, the following null hypothesis suggests itself:

Null Hypothesis 4.6. *When examining the time it takes a subject to encode an opinion ($encodeTime$), there is a decrease across $initialPresentation$ from YES to NO, when $encodeType$ is SYMBOL.*

For words, a priming effect would be expected, because presumably there no syntactic barriers to the subject responding to the task to given an opinion when the choices involve words. Therefore, the following null hypothesis seems appropriate:

Null Hypothesis 4.7. *When examining the time it takes a subject to encode an opinion ($encodeTime$), there is no decrease across $initialPresentation$ from YES to NO when $encodeType$ is WORD.*

The results relating to Null Hypotheses 4.6 and 4.7 are shown graphically in Figure 4.16. As shown, the priming effects are substantial. For Null Hypothesis 4.6, $p < 1.0\,(df = 3128)$, and so cannot be rejected. For Null Hypothesis 4.7, $p < 0.001\,(df = 3128)$, and so can be rejected. It is interesting that on average, the encoding time for symbols is less than that of words for both initial and secondary topic presentation. Moreover, the apparent priming benefit is larger when symbols are presented first.

4.4.4 Encoding Learning Effects

As mentioned in Section 4.3.3 under Task 6, ten topics were used and each was presented twice. The sequence was randomized, so there are 0-9 potential previous exposures to word choices or symbol choices. It seems plausible that subjects

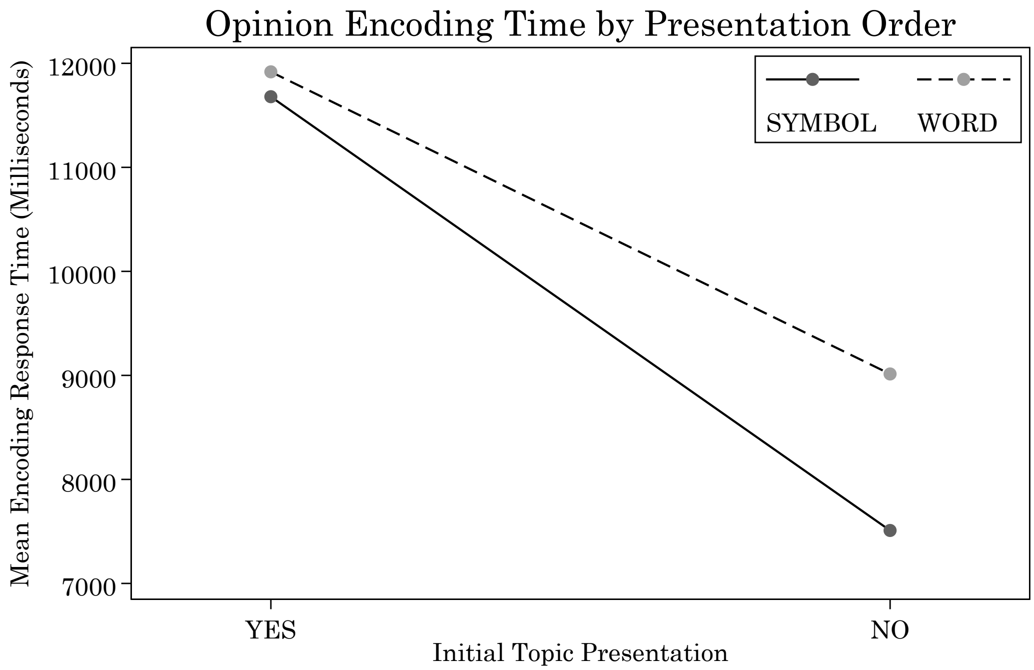


Figure 4.16: *encodeTime* by *initialPresentation* (YES or NO) and *encodeType* (WORD or SYMBOL)

would improve their rate of response as they progress through Task 6—demonstrating a learning effect, which is measurable. From a theoretical perspective, it would make sense that since symbolic opinion encoding schemes are new, and words are not, that learning effects would be more pronounced in the former than the latter. The graph shown in Figure 4.17 shows the sequential relationship (by number of previous encoding attempts) between average encoding time for words and symbols. If there are zero or small learning effects, then subject response times would not show much of a decrease from the first exposure to the last.

With Figure 4.17 as a guide, there appear to be two distinct phases of subject experience (*experientialPhase*), LEARNING—from 0-2 *previousAttempts*, and APPLYING, from 3-9 *previousAttempts*. Continuing the theoretical linkage discussed above, it would make sense if learning effects were greater when using symbols than when using words. The following null hypothesis allows us to test this proposition:

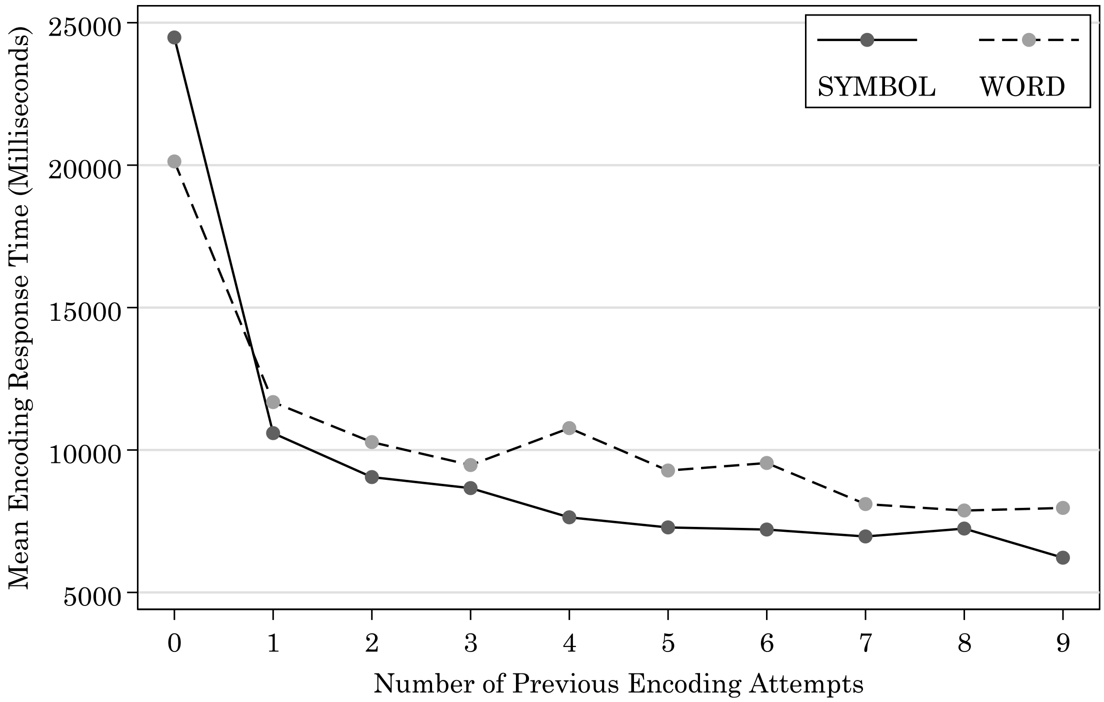


Figure 4.17: Mean *encodeTime* by *previousAttempts* and *encodeType* (SYMBOL and WORD).

Null Hypothesis 4.8. *The decrease in opinion encodeTime across experientialPhase values LEARNING to APPLYING should be the same for encodeType values of WORD and SYMBOL.*

The graph shown in Figure 4.18 shows the relationship between the variables in Null Hypothesis 4.8. There is a more substantial learning effect for symbol encoding than word encoding $p < .001 (df = 624)$, so we can reject Null Hypothesis 4.8.

4.4.5 Encoding Topic Category Effects

As discussed previously, ten *topic* values were defined, with each topic belonging to one of four *topicCategory* values. An interesting question is whether or not the *encodeTime* value would differ across symbolic encoding tasks by *topicCategory*

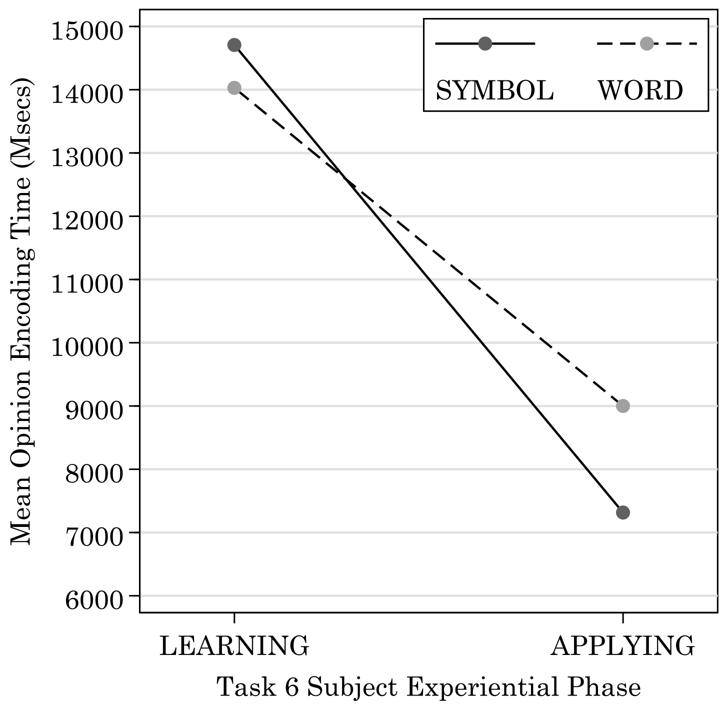


Figure 4.18: Mean *encodeTime* by *experientialPhase* (LEARNING and APPLYING) and *encodeType* (WORD and SYMBOL)

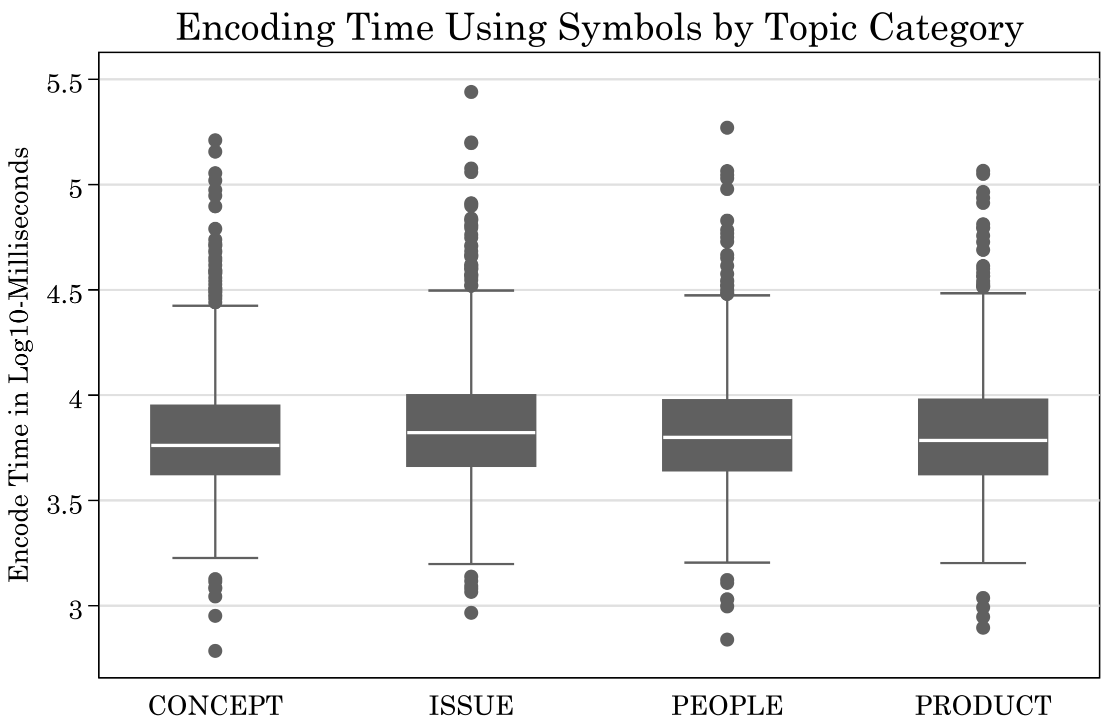


Figure 4.19: Log10 distribution of symbolic $encodeTime$ by $topicCategory$

. A large disparity for a particular topic category value may be an indication that symbolic representations of sentiment are not a good fit for that type of topic.

Null Hypothesis 4.9. *The opinion* $encodeTime$ *for* $encodeType$ *value SYMBOL, is not influenced by the* $topicCategory$ *to which the target belongs.*

As shown in Figure 4.19, there is very little disparity across $topicCategory$ values. Using ANOVA to look at the relationship between $encodeTime$ and $topicCategory$ values for symbolic encoding tasks shows detectable difference in the variances by topic category with $p < .1\,(F = 2.46)$. We can reject Null Hypothesis 4.9 at a weak level of significance.

While there might not be much variation between topic categories, there may be some variation by encoding scheme ($scheme$). If there are interactions between $scheme$ and $topicCategory$, then that might be an indication that some symbolic

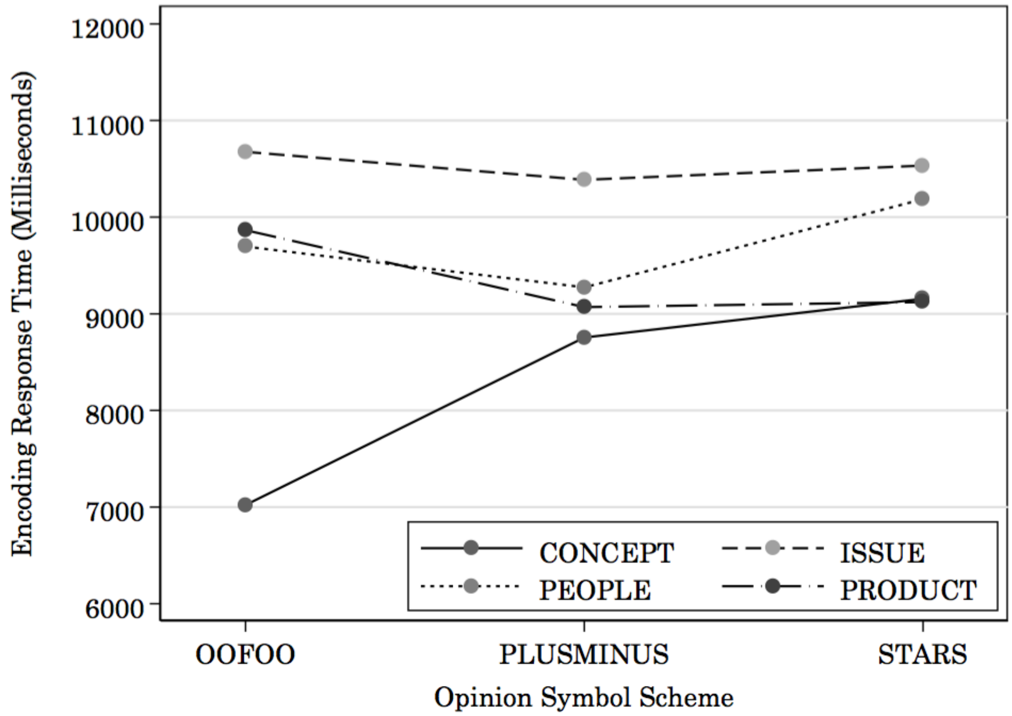


Figure 4.20: Mean *encodeTime* by *topicCategory* by encoding *scheme* (OOFOO, PLUSMINUS, and STARS)

representations of opinion are better suited to certain types of topics than others. The following null hypothesis reflects this relationship:

Null Hypothesis 4.10. *The opinion encodeTime for topics belonging to a particular topicCategory will not vary by the symbolic opinion encoding scheme used.*

The graph representing the relationship defined in Null Hypothesis 4.10 is shown in Figure 4.20. Using ANOVA to examine the relationship proposed in Null Hypothesis 4.10 results in not being able to reject the null hypothesis, with $p < 0.65$ $(F = 0.75)$). While there appears to be a differential response time for *scheme* value OOFOO for *topicCategory* CONCEPT, the variance of the data overwhelms the central tendencies—making it unfruitful to generalize.

4.4.6 Decoding Rates

In Task 7, the subject was asked to decode someone else's opinion expressed in symbolic form, choosing from among polar opposite values across five dimensions of opinion (See Figure 4.8). Research Question 4.2 seeks to explore demographic effects. Exposure to cryptic symbolic notations is a part of the social media experience. Therefore, it makes sense to expand the question about decoding response times to include an addition dimension: $socialMediaUsageLevel$ (LIGHT, MODERATE, or HEAVY). If the opinion decoding process is difficult for subjects due to the symbology, it would make sense that the decoding time for subjects who are HEAVY users of social media would be shorter than those for LIGHT users. The following null hypothesis explores this relationship:

Null Hypothesis 4.11. *The opinion $decodeTime$ will not vary by $socialMediaUsageLevel$; specifically, LIGHT users of social media will not be greater than that of HEAVY social media users.*

As shown in Figure 4.21, the results are the opposite of what was expected, with $p < .99$ $(df = 2429)$ —not allowing Null Hypothesis 4.11 to be rejected. It turns out that LIGHT social media users decode opinions expressed using a structured sentiment scheme faster than HEAVY users, with $p < .015$ $(df = 2429)$.

4.4.7 Decoding Learning Effects

Just as with encoding learning effects discussed in Section 4.4.4, it can be expected that if subjects are genuinely taking part in the experiment—learning would occur and the time to decode ($decodeTime$) would decrease from the 1st attempt to the 15th. Figure 4.22 shows the relationship between number of attempts, $decodeTime$, $decodeAttempts$, and $socialMediaUsageLevel$.

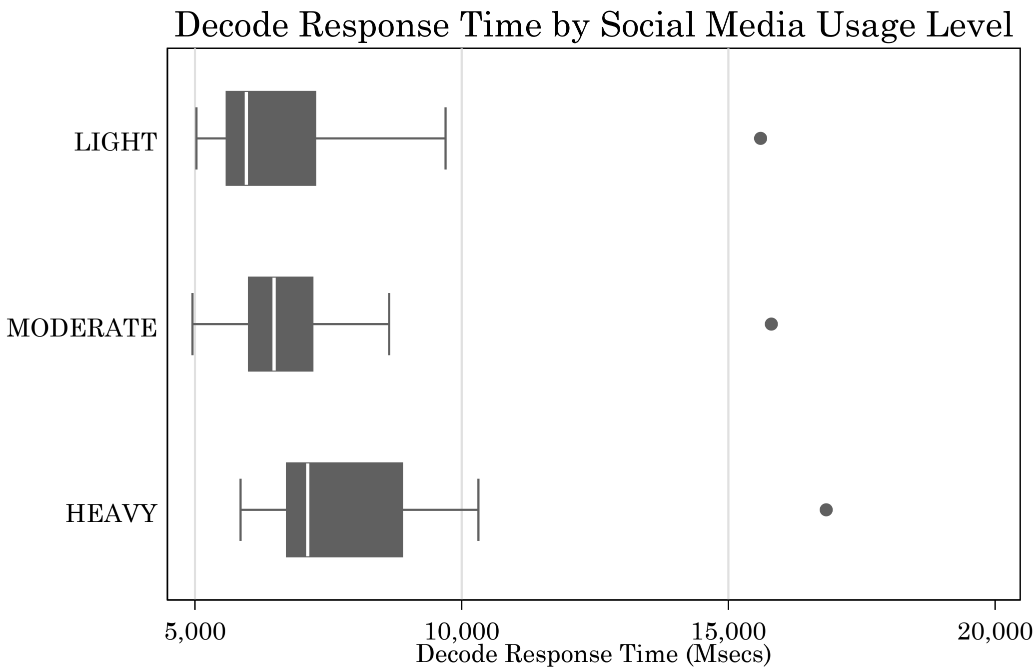


Figure 4.21: Distribution of *dedcodeTime* by *socialMediaUsageLevel* (LIGHT, MODERATE, and HEAVY)

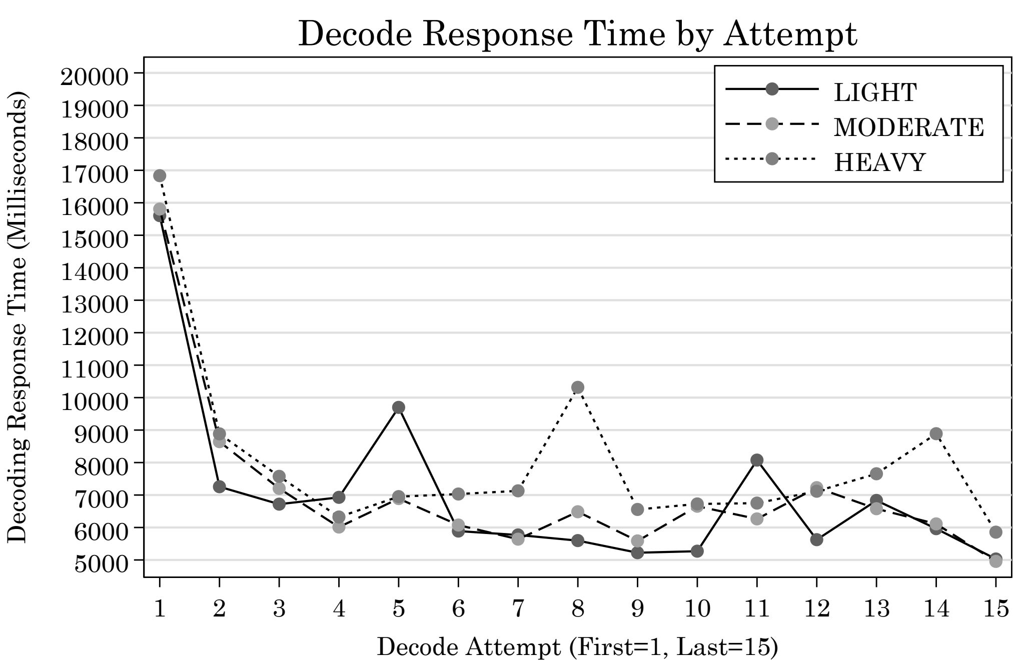


Figure 4.22: Mean *dedcodeTime* by *decodeAttempts* and *socialMedia UsageLevel* (LIGHT, MODERATE, and HEAVY)

Null Hypothesis 4.12. *The opinion $decodeTime$ will not vary by $socialMediaUsageLevel$ or $experiantialPhase$ (LEARNING=0-2, and APPLYING=3+.)*

An ANOVA on the dependent variable $decodeTime$ using the two independent variables from Null Hypothesis 4.12 individually as well as the interaction between the two yielded the following results. The variable $socialMediaUsageLevel$ was not a useful indicator of $decodeTime$, $p < .12$ $(F = 2.19)$, neither was the combination of $socialMediaUsageLevel$ and $experiantialPhase$, $p < 0.7$ $(F = 0.4)$. However, the mean $decodeTime$ was significantly different across $experiantialPhase$ values of LEARNING ($mean = 12086$, $sd = 10638$) and APPLYING ($mean = 6796$, $sd = 11237$), $p < 0.001$ $(F = 161)$.

4.4.8 Decoding Consistency

As discussed in Section 4.3.3 under Task 7, subjects were presented with 15 decoding subtasks—one opinion of another person in each. Of the 15 decoding subtasks, 10 are random selections. The remaining five decoding subtasks are re-presentations of the subjects encoded opinions from Task 6. The purpose of this design in S3 was to allow for a correspondence analysis between the subject's own opinions encoded with a particular $scheme$, and the subjects later re-interpretation of those same symbolic encoded opinions as expressed by another person.

For an example of the conditions required for a match to occur, see Task 6, Figure 4.6 for an original symbolic encoding of how the subject feels about COKE: **`#coke+++`**. Later in Task 6 (Figure 4.7) the subject was again asked about COKE, but given a randomly ordered list of $opinionWords$ to choose from. In the example from Task 6, the subject chose LIKE for the target COKE.

Continuing with the example, in Task 7, the subject is presented with the same symbolic sentiment that he or she selected in Figure 4.6, **`#coke+++`**. The subject is then asked to choose from among a randomly ordered list of $opinionWords$. A

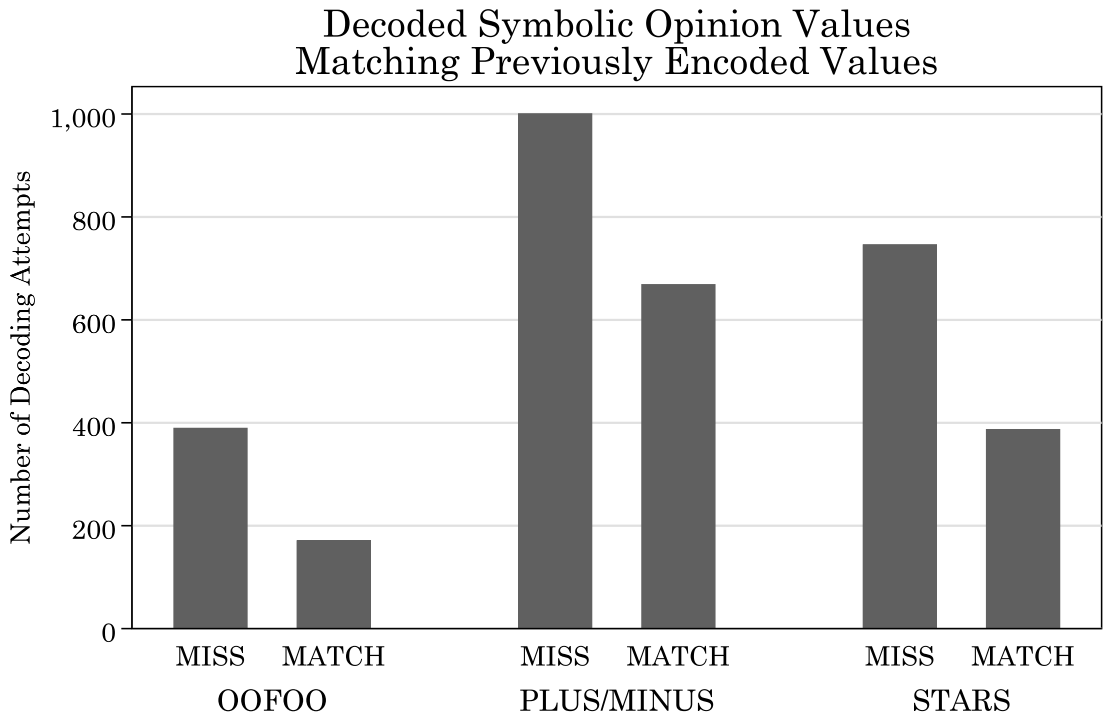


Figure 4.23: Matching Task 7 decoding and Task 6 encoding by encoding *scheme*

match is recorded if the *opinionWords* selected in Figure 4.8, in this case LIKE, match the word selected in Figure 4.7, LIKE.

The null hypothesis, then, to look at this relationship can be defined as follows:

Null Hypothesis 4.13. *The opinionWords selected by a subject when encoding a symbolic opinion for a target in Task 6, will not match the opinionWords selected by the same subject when decoding the same symbolic opinion for the same target in Task 7, beyond what can be expected through random selection.*

As shown in Figure 4.23, the decode match rate appears to be well above chance for each symbolic encoding scheme—so Null Hypothesis 4.13 can be rejected. To confirm the effect, a proportion test was used with the threshold for randomness being 1:10, because *opinionWords* has 10 values. In the case of each *scheme* (OOFOO, PLUS/MINUS, and STARS), $p < 0.001$ $(obs = 561, 1670, 1133)$.

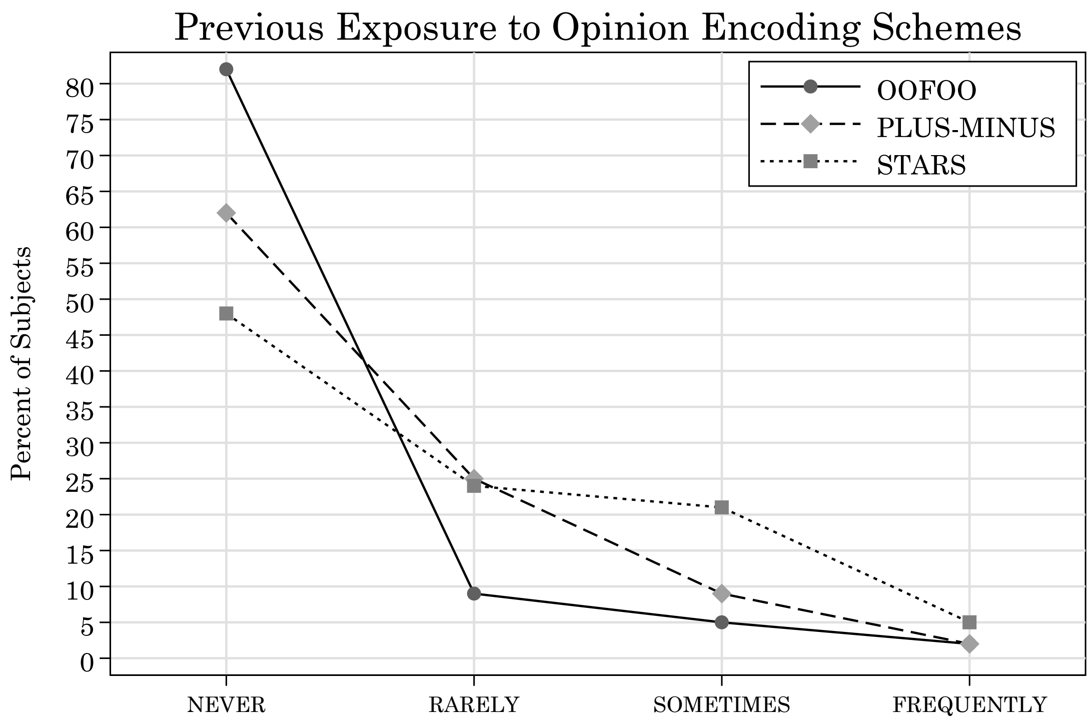


Figure 4.24: Subject $exposureFrequency$ by encoding $scheme$

4.4.9 Decoding Experience

Having completed the 20 encoding attempts in Task 6 and the 15 decoding attempts in Task 7, subjects are asked in Task 8 to answer three (3) experience and usage questions. The first, question seeks an understanding of the level of exposure to opinion encoding schemes like OOFOO, PLUS-MINUS, and STARS. Not surprisingly, as shown in Figure 4.24, the reported exposure levels are very low. These encoding schemes are only a curiosity at this point within social media, lacking any organizing force to give an impetus for adoption. The reported values in Figure 4.24 impute some veracity to the other findings in this paper, but no testable hypothesis presents itself.

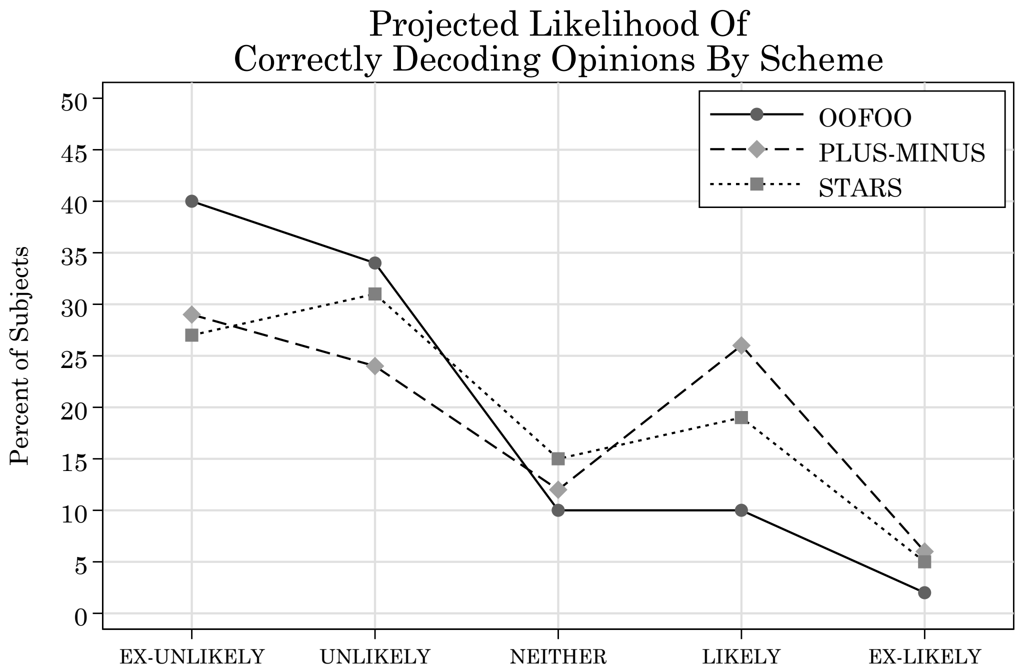


Figure 4.25: Subject projected likelihood of correctly decoding symbolic opinion (*likelihoodDecoding*) by encoding *scheme*

4.4.10 Decoding Prospective Proficiency

The second question in Task 8 asks the subject to project the likelihood that he or she would be able to accurately identify an opinion written using each encoding *scheme* . The results shown in Figure 4.25 show a level of confidence, but the UNLIKELY to LIKELY ratio is approximately 3:1 (185 to 68).

4.4.11 Encoding Prospective Proficiency

The last question in Task 8 asks the subject to project the likelihood that he or she would be able to correctly encode an opinion in the future using a symbolic encoding scheme. The results shown in Figure 4.26 show no measurable confidence. The UNLIKELY to LIKELY ratio is approximately 15:1 (251 to 15).

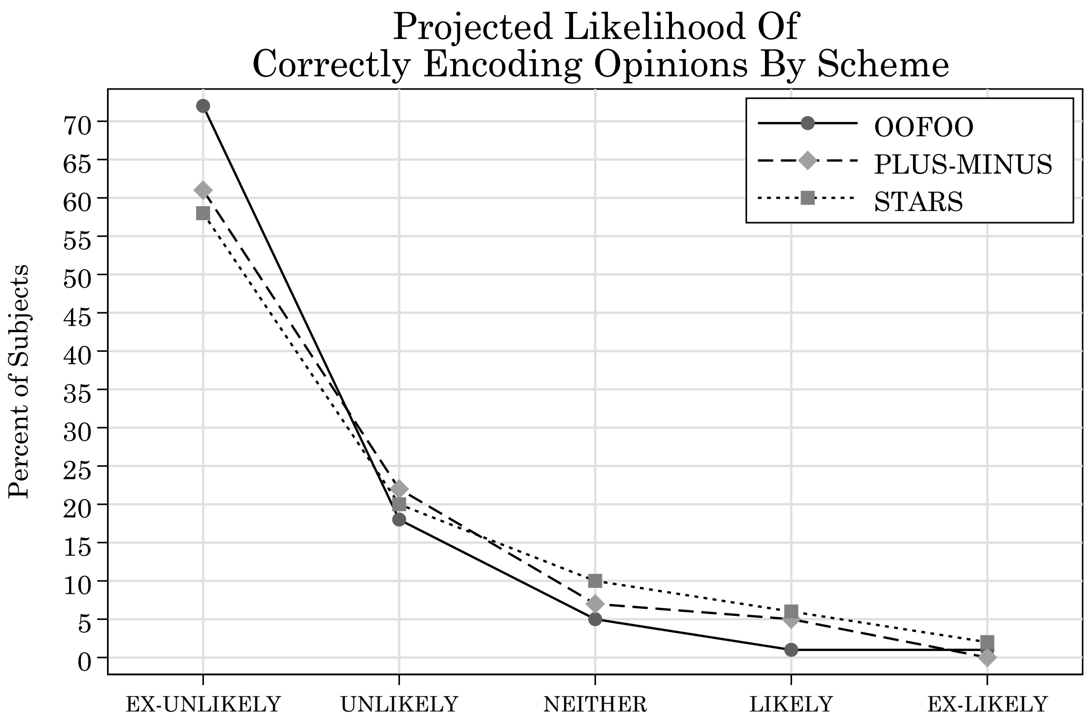


Figure 4.26: Subject projected likelihood of correctly encoding symbolic opinion (*likelihoodEncoding*) by encoding *scheme*

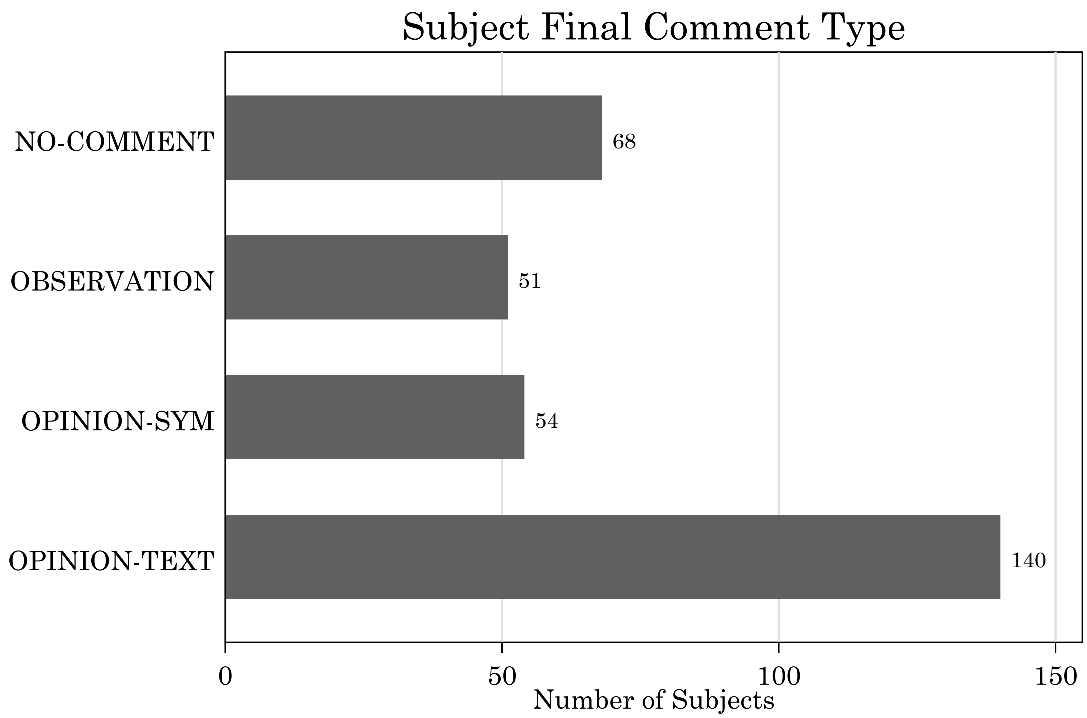


Figure 4.27: Subject final comment intention and symbolic encoding proficiency

4.4.12 Actual Encoding Proficiency

The final task subjects were asked to complete, Task 9, was to "write some opinions you have" using either conventional prose or a symbolic opinion encoding scheme. A helpful guide was provided above the text box, explaining the relative values of the different encoding schemes. The variable values captured in Task 9 by S3 are shown in Table 4.6. As shown in Figure 4.27, the ratio of $subjectCommentType$ values of OPINION-TEXT to OPINION-SYM was approximately 3:1 (140 to 54), with 16.3% of subjects (51) making an objective statement and 21.7% of subjects not providing any response to Task 9.

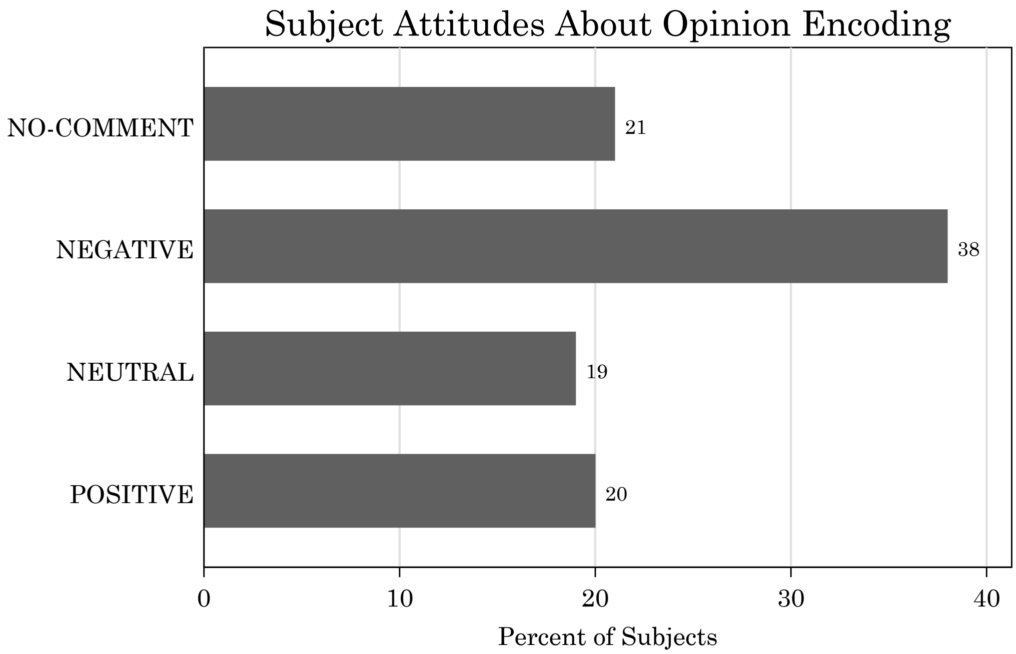


Figure 4.28: Subject attitudes about using opinion encoding schemes.

4.4.13 Symbolic Opinion Attitudes

The comments made by subjects in Task 9 were reviewed by the primary investigator for expressions of sentiment toward using opinion encoding schemes. A variable *subjectCommentBias* was assigned a value of NO-COMMENT, NEGATIVE, NEUTRAL, or POSITIVE—and the distribution of these values across the subject population is shown in Figure 4.28.

Examples of some of the subject comments marked NEGATIVE are shown below.

- These notations are idiotic ans asinine! I would only expect to see such syntax from persons who did not complete high school ... [this] applies to all of these symbols, they and the person using them must be an idiot.

- I don't think using these sentiment notations are that effective... it's better to just use regular text.

- I am confused.

- Why? For the love of god, WHY? What's wrong with using words? Makes so much more sense. I'm scared that in the near future people will actually use 'oooo' and 'fooo' in conversations to describe their opinions. Or actually say 'plus plus plus plus' when describing how awesome the food is. I stay away from Twitter and I sure damn hope this shit isn't real.

- These symbols are ignorant and proof that people are becoming less educated when they have more resources to become knowledgeable. People who use these symbols are the filth of the evolving world.

Examples of some of the subject comments marked POSITIVE are shown below.

- #summer++++ foo#school #puppies*****

- I have to say, #Coke+++ I have to say, #Deathpenalty+++ I have to say, #Family***

- ooo#food ooo#summer+++ foo#homework— foo#exams

- I totally voted for #Obama++ but really only because #Romney— was such a twat. I love me some oooo#Merica! Screw them foo#arabs! I really like extra credit #surveys*** because #failing** sucks.

- oooo#Movies #Oatmeal++++

- Honestly, I've never seen these notations used on social media for anything, but just from common sense I guess what each of them meant.

- I have to say, ooo#this

4.5 Conclusions

Structured opinion encoding schemes are only postulates—not concrete phenomena for which there is a body of evidence. The naive, negative, and bewildered attitudes that subjects expressed in Task 9 and shown in Figures 4.24 to 4.26

and 4.28 are clear indications of this. Moreover, the results presented in this paper cover a wide variety of phenomena, examined through a single data set—for which there does not appear to be any precedent. Conclusion, therefore, is too strong a word. However, a number of interesting outcomes merit review.

The asymmetry in opinion read vs. write frequency shown in Figure 4.11 is supported by the pragmatics of social media—however, seeing the 8:1 ratio of opinion WRITERS to READERS was still surprising. The large number of subjects who generate hundreds of opinions per week while consuming few or none was significant (Figure 4.12). The practical implications are straight-forward: lots of opinions are being expressed, but few are consuming them. Of course, that is the point of structured sentiment encoding—an attempt to "enfranchise" the long-tail of social media content generators by making all those opinions reliably decodable. The results of Task 8 and Task 9 demonstrate that the barriers to the adoption of structured sentiment encoding schemes may be numerous. However, the cognitive process of performing the encoding does not seem to be one of them.

The encoding rate advantage of symbols over words shown in Figure 4.13 and the gender-neutral character of encoding rates in Figure 4.14 gives some indication that if a suitably descriptive encoding scheme was identified, the cognitive load associated would not necessarily pose a barrier to adoption. Especially in light of the finding shown in Figure 4.15 that social media usage level was not a driving factor in a subject's ability to encode. Also, the priming effects present in Figure 4.16 and the learning effects apparent in and Figure 4.17 and Figure 4.18 demonstrate that a first-time user of structured sentiment can rapidly assimilate a syntax and how to use it. Target categories (Figure 4.19) or the scheme used (Figure 4.20) did not appear to be influential in the subjects abilities to encode opinions.

Decoding rates showed similar effects to those shown in encoding, though decoding rates were the opposite of what might logically be expected. Figure 4.21

showed that HEAVY users of social media took longer to decode opinions than LIGHT users. Numerous explanations are possible—such as HEAVY users have more analogs to wade through, and therefore will take longer. It is an interesting finding, but its meaning is unclear. The decoding learning effects were equally pronounced to those of encoding (Figure 4.22). Certainly an interesting finding is that despite the protestations shown in Figure 4.23, subjects provided matching decoding and encoding responses to Task 6 and Task 7 at levels well above chance.

Lastly, the responses from subjects when given the opportunity to express themselves openly, were passionately negative in many cases. The use of symbolic representations of sentiment triggered a backlash against what might be characterized as the cryptic and sophomoric tendency in social media to do things like **`ooo#pizza`**. There was a substantial plea for the use of conventional prose—which can easily be understood by humans: *"I think it's stupid and pointless. USE YOUR WORDS."* Nevertheless, there seems to be some merit in the use of symbols in expressing opinion. The 51 affirmative usages of a hypothetical symbolic opinion encoding notation (Figures 4.27 and 4.28) are a testament that language itself is flexible—and the propensity to invent by some stimulates experimentation by others.

Looking ahead, the development and adoption of a structured sentiment representation may continue in the shadows of social media for some time, until the "need to say something" is eclipsed by the "need to to be heard."

Recommendations

The vision of structured sentiment is unambiguous opinion encoding, enabling precise opinion decoding. This exploratory paper introduces the study of structured sentiment. The actual symbols, their relationship to the target, and alternative notations were not pursued—but will need to be if a credible syntax is to be developed. It may be that some of the elements of the execution of this study can be re-used, but the brightest and whitest space for inquiry is likely to be around the definition of encoding schemes (like UVML) which raise the value of opinion expression in social media.

Acknowledgements

Special thanks are due to professors Michael Dahlstrom (JLMC 101) and Jay Newell (ADVRT 230) for their encouragement and support in the execution of this study. Both were interested, engaged, and enthusiastic about the subject matter of this study, and their energy no doubt inspired many given the high levels of participation.

BIBLIOGRAPHY

Aoki, S. & Uchida, O. (2011). A method for automatically generating the emotional vectors of emoticons using weblog articles. In *Proc. 10th WSEAS Int. Conf. on Applied Computer and Applied Computational Science, Stevens Point, Wisconsin, USA*, (pp. 132–136).

Baker, C., Fillmore, C., & Lowe, J. (1998). The berkeley framenet project. In *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics-Volume 1*, (pp. 86–90). Association for Computational Linguistics.

Berners-Lee, T., Hendler, J., Lassila, O., et al. (2001). The semantic web. *Scientific american*, *284*(5), 28–37.

Bizau, A., Rusu, D., & Mladeni, D. (2011). Expressing opinion diversity.

Cover, T. M. & Thomas, J. A. (1991). Entropy, relative entropy and mutual information. *Elements of Information Theory*, 12–49.

Das, S. & Chen, M. (2001). Yahoo! for amazon: Extracting market sentiment from stock message boards. In *Proceedings of the Asia Pacific finance association annual conference (APFA)*, volume 35, (pp.43). Bangkok, Thailand.

Duggan, M. & Brenner, J. (2013). *The demographics of social media users, 2012*, volume 14. Pew Research Center's Internet & American Life Project.

Go, A., Huang, L., & Bhayani, R. (2009). Twitter sentiment analysis. *Entropy*, *17*.

Goldfarb, C. & Rubinsky, Y. (1990). *The SGML handbook*.

Hancock, J., Thom-Santelli, J., & Ritchie, T. (2004). Deception and design: The impact of communication technology on lying behavior. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, (pp. 129–134). ACM.

Ku, L., Liang, Y., & Chen, H. (2006). Opinion extraction, summarization and tracking in news and blog corpora. In *Proceedings of AAAI-2006 Spring Symposium on Computational Approaches to Analyzing Weblogs*, (pp. 100–107).

Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis Lectures on Human Language Technologies*, *5*(1), 1–167.

Macdonald, C., Ounis, I., & Soboroff, I. (2007). Overview of the TREC 2007 blog track. In *Proceedings of TREC 2007*.

Nasukawa, T. & Yi, J. (2003). Sentiment analysis: Capturing favorability using natural language processing. In *Proceedings of the 2nd international conference on Knowledge capture*, (pp. 70–77). ACM New York, NY, USA.

Pang, B. & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, *2*(1-2), 1–135.

Pang, B., Lee, L., & Vaithyanathan, S. (2002). Thumbs up?: sentiment classification using machine learning techniques. In *Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10*, (pp. 79–86). Association for Computational Linguistics.

Phillips, E. (2011a). Toward a universal sentiment taxonomy: A multi-language analysis of sentiment traces in a large blog corpus.

Phillips, E. (2011b). The universal voting markup language (uvml). In *IETF, DRAFT RFC*.

Phillips, E. (2015). Speq-ing the truth: States, processes, effects, and quality (speq) model for opinion mining and sentiment analysis.

Post, R. C. (1990). The constitutional concept of public discourse: outrageous opinion, democratic deliberation, and hustler magazine v. falwell. *Harvard Law Review*, 601–686.

Rainie, L. (2006). Blogger callback survey. Technical report, Princeton Survey Research Associates, Pew Internet & American Life Project.

Read, J. (2005). Using emoticons to reduce dependency in machine learning techniques for sentiment classification. In *Proceedings of the ACL Student Research Workshop*, (pp. 43–48). Association for Computational Linguistics.

Simmons, K. (2014). Emerging nations embrace internet, mobile technology. *Washington, DC, USA: Pew Research Center*.

Tsur, O. & Rappoport, A. (2012). What's in a hashtag?: content based prediction of the spread of ideas in microblogging communities. In *Proceedings of the fifth ACM international conference on Web search and data mining*, (pp. 643–652). ACM.

Turney, P. D. (2002). Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. In *Proceedings of the 40th annual meeting on association for computational linguistics*, (pp. 417–424). Association for Computational Linguistics.

Yi, J., Nasukawa, T., Bunescu, R., & Niblack, W. (2003). Sentiment analyzer: Extracting sentiments about a given topic using natural language processing techniques. In *Third IEEE International Conference on Data Mining, 2003. ICDM 2003*, (pp. 427–434).

CHAPTER 5. GENERAL CONCLUSIONS

The following comments are limited to those remarks which are not otherwise covered in Chapters 2 to 4.

5.1 General Discussion

The three lines of inquiry presented in Sections 1.1 and 1.2 provide a framework for the three important and related research outcomes derived from the research supporting this dissertation.

Is opinion expression universal? If so, how?

Chapter 2 applied scholarship from multiple disciplines to derive a taxonomy of sentiment types, and then evaluated the veracity of that taxonomy against a large corpus of 400M+ social media documents in 15 languages. The results were compelling, and this doctoral candidate is comfortable asserting that opinion is indeed universal. This assertion is derived more from the demonstrated portability and universality of adverbial exemplars than the particular taxonomy developed in Chapter 2 through the analysis of social science scholarship. It certainly seems plausible that a better or different taxonomy of sentiment types may be developed through other means; however, the character of that taxonomy would likely align with the Likert-ness of U18. In that sense, the research supporting the first paper is an archetype for the development of a sentiment type taxonomy more than the presentation of a definitive sentiment type taxonomy.

Is opinion processing universal? If so, how?

The literature from the field of opinion mining and sentiment analysis shows a lack of a detailed theoretical model of opinion—one which spans the life of an opinion from private state to summarization. The second paper, presented in Chapter 3, follows a tedious but systematic method of qualitative analysis (QMS) to an exhilarating and clarifying outcome: the States, Processes, Effects, and Quality (SPEQ) model for opinion mining and sentiment analysis. SPEQ includes seven (7) distinct states of opinion, six (6) process which govern transitions between those states, and five (5) measures of quality (including 3 measures of integrity) for those processes. The "chain of custody" which SPEQ defines has the potential to improve the veracity of future work in opinion mining and sentiment analysis.

Is there a way to leverage both the universality of opinion expression and the universality of opinion processing to create a more reliable form of encoding and decoding opinions? If so, how?

In the spirit of the first two papers, the third paper also reaches beyond conventional notions of opinion mining and sentiment analysis research. Relying on the universality of opinion in social media demonstrated in Chapter 2 and anchoring key concepts in the SPEQ model developed in Chapter 3, structured opinion encoding schemes were postulated and investigated. As mentioned in Chapter 2, structured opinion encoding schemes are "only postulates—not concrete phenomena for which there is a body of evidence." These postulated opinion encoding schemes, however, are practical instantiations of the conceptual link between the SPEQ states of RECORDED and INTERPRETED opinion. SPEQ also explains the 8:1

WRITERS to READERS ratio in the context of current opinion mining and sentiment analysis. Methods of recording opinion are abundant; however, a method for precisely encoding opinion has not yet been defined; therefore, no mechanism exists to decode opinion reliably.

The experiment in Chapter 4 introduced subjects to a structured opinion encoding syntax. An important finding was derived from the learning and priming effects shown in Figures 4.17 and 4.18 for encoding. Likewise, learning effects for decoding were also shown in Figure 4.22. The implications are that the cognitive load associated with processing symbolic representation of opinion may be on par with those of conventionally written text. Perhaps the most interesting finding was that round-trip encoding/decoding correspondence rates show a semantic stickiness to opinions encoded with symbols—even in subjects who declare an aversion of such symbology. Therefore, as stated in Chapter 4, the development and adoption of a structured sentiment representation may continue in the shadows of social media for some time, until the "need to say something" is eclipsed by the "need to to be heard" coincident with a platform to enable the hearing.

5.2 Recommendations for Future Research

A number of specific recommendations for future research have been presented in each of the three papers in this dissertation; however, an important "next step" has not been otherwise directly mentioned. As more and more opinion content becomes intelligible, another layer of collaborative analysis becomes an imperative. As expressed by Fishkin et al. (2008):

> "If one just invites the public to open town meetings, the appearance of mass participation may belie practices in which organized interests actually dominate . . . organization is an unequally distributed resource and open forums can be captured through efforts at mobilization." (p. 1)

Innovative and important work is being done on the concept of deliberative voting (Luskin, Fishkin & Plane, 1999; Fishkin, Luskin & Jowell, 2000; Fishkin, 2000; Fishkin & Luskin, 2005; Fishkin, He, Luskin & Siu, 2006; Fishkin, He & Siu, 2008). This scholarship could be integrated with reputation systems research and the concepts developed in this dissertation. The result may enable the kind of effective decision-making envisioned by Fishkin and Luskin, but on a mass scale.

Authorship

The following attestation confirms the authorship of all portions of this dissertation, supporting artifacts, and related assets. The listed author is the sole author of this dissertation. The ideas and works of others have been acknowledged and referenced, whether published or unpublished. All non-academic support of this work from within this university and any support whatsoever from outside this university has been acknowledged. This dissertation does not contain work extracted from a thesis, dissertation or research paper previously presented for another degree or diploma at this or any other university.

APPENDIX A. CITESCAN: A BIBLIOGRAPHY, CITATION, AND CONTENT ANALYSIS AND QUERY TOOL

This appendix provides some implementation details regarding the CiteScan utility developed in support of this research. CiteScan was developed specifically to facilitate the analysis of large numbers of published research papers given the diversity of materials involved in this research. No suitable (and affordable) tools could be identified, so some time and effort was invested in creating CiteScan. This work, though ancillary to the core concerns of this research, is provided here because CiteScan constitutes an innovative approach to analysis and synthesis of large amounts of published research. If formalized and implemented on a larger scale, CiteScan may improve the efficiency and quality of literature reviews more generally.

A.1 Capabilities

CiteScan was developed to answer the following types of questions regarding a corpus of research papers:

1. Given term T, what papers are most focused on T?
2. Given paper A, which papers were most relied upon in A?
3. Given term T and proximity term U, what papers discuss T in the presence of U?

4. Given paper A, which authors were most relied in A?

The CiteScan workflow reflects the nature of knowledge acquisition, but more specifically parallels the use of corpora to learn. Brill (1993) used a small annotated corpus and a large unannotated corpus to develop a series of transformations capable of annotating new corpora. To the use of corpus linguistics and corpora as a learning environment, CiteScan adds capabilities which enable the process of concept explication outlined in Chaffee (1991) [diagram on pp. 6]: identify concepts, review literature, analyze meanings, develop new definitions, and repeat. The Chaffee (1991) workflow is described as iterative, with many paths back to review and refine earlier learnings. Similarly, CiteScan enables the user to use corpus linguistics techniques, namely learning the meaning of core terms and their related contextual cues through an interactive and exploratory process akin to concept explication.

A.2 Workflow

CiteScan uses a 4-step process shown in Figure A.1 to index document content and build a searchable database. The 5th step is to use the query facility to explore as described above. The data structures which enable the query capabilities are shown in Figure A.2.

Step 1 : Preprocess PDF Document

In Step 1, the PDF document is converted to text and errant and unusual Unicode values are converted to normalized values (finding a 'more familiar' code point). Unicode validation is required because the reliability of pdf to text decoding did cause some problems with name recognition.

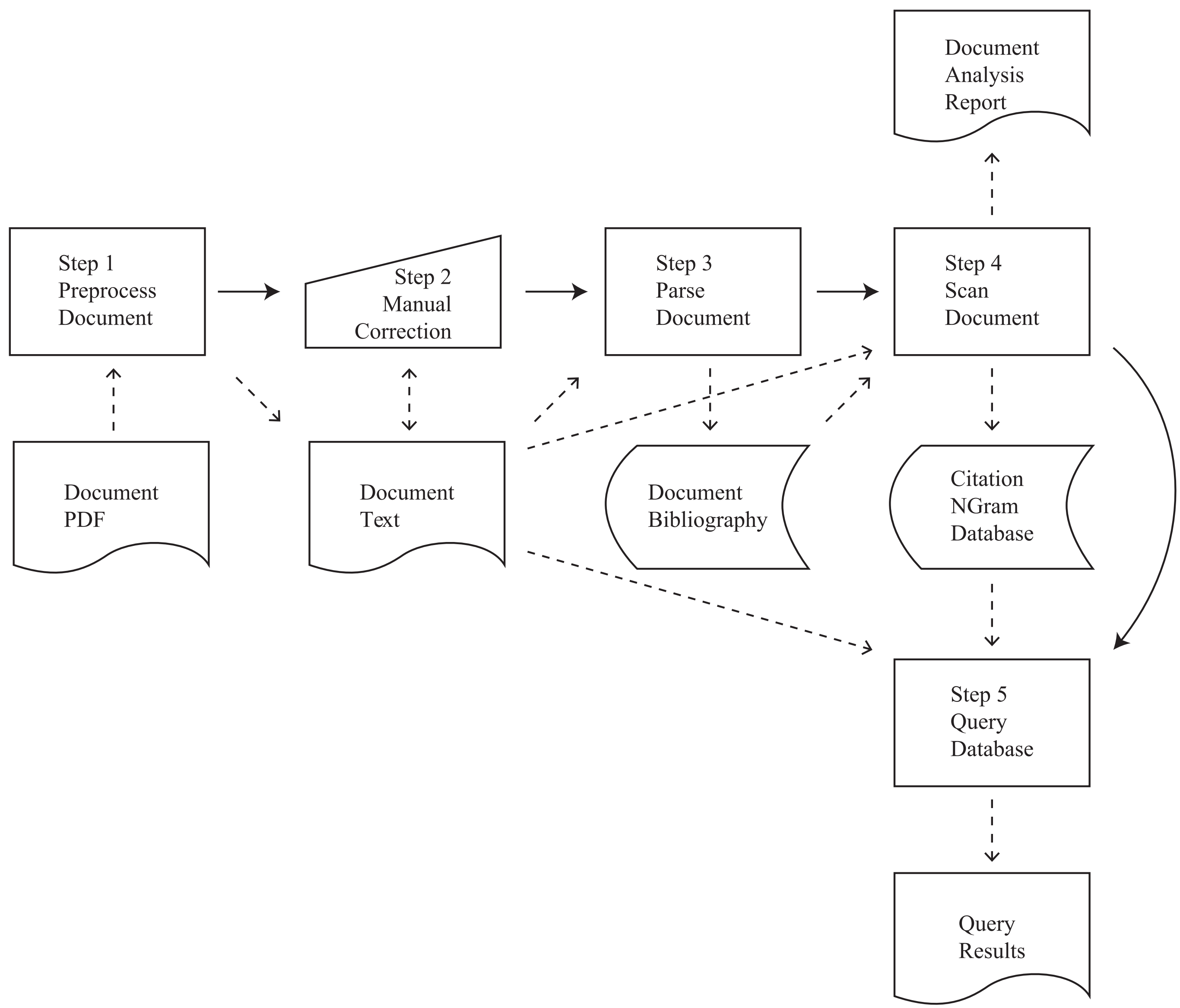


Figure A.1: CiteScan tool workflow

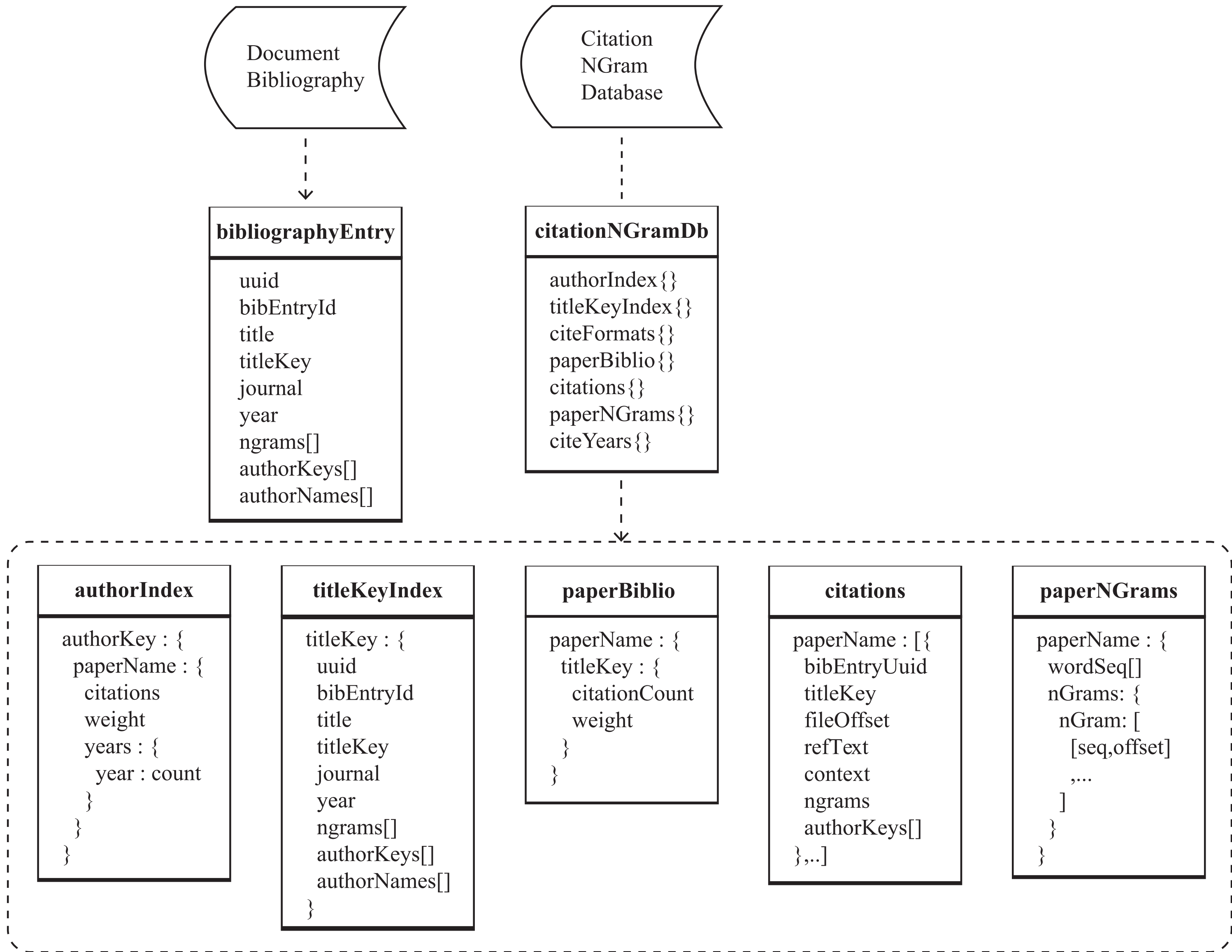


Figure A.2: CiteScan tool data model

Step 2 : Manual Correction

In Step 2, the text version of the source PDF is modified by hand. There are only a few classes of manual modifications which proved to be necessary:

1. Bibliographical pathology, i.e., not enough information or incorrect information was included in the bibliography, such as “Roberts, Charles, F., Robb, A. 1990.”

2. Citation pathology, i.e., not enough information or incorrect information was included in a paper’s citation of a bibliographical reference, such as “Fuches (2010)” when there is not 2010 article associated with this author.

3. Citation ambiguity, i.e., confusing references to citations, such as “Plank (2006-2010).”

These errors were surprisingly uncommon. Most papers were fully parsable after less than 5 minutes of manual correction and 2-3 attempts to parse.

Step 3 : Parse Document

The parsing step builds a bibliographical parser based upon the BNF specification for the citation style used in the document. The BNF syntax used is a modified BNF grammar, which supports some macro expansion. Table A.1 shows the BNF expression for the type of bibliographical entry which contains a JOURNAL reference.

The output of the parse step is a JSON file which contains the normalized bibliography entries found in the references or bibliography section of the paper or papers found in the document. The JSON file is created by rendering the BNF parse tree of the bibliography entries found in the document using a customized tag-aware code generator written in NODE.JS.

Step 4 : Scan Document

In the scan step, the document-level JSON file containing the bibliography entries are integrated into the global author and n-gram indexes. Then, citation scanning regexes are generated for each bibliographical entry using the citation format the document uses. Table A.2 shows the regexes generated for "[3]" and variants as found in an IEEE formatted paper.

APA style references require a more complex set of regular expressions, as shown below in Table A.3.

Additionally, the lexemes found in the conformed text file from the prepare step are indexed, to allow for proximity searches. The scan step also produces a report for the document, listing the most influential papers referenced in the document, the most influential authors whose works are cited in the document, along with most frequently used unigram, bigrams, and trigrams. Examples are shown in Tables A.4 and A.5.

Step 5 : Query Document

The query facility is crude, but does an adequate job of extracting relevant text. Table A.6 shows an example of the query output.

Table A.1: BNF grammar specification for an IEEE bibliographical entry with the year, paper title, and journal title.

Level 0 seq	paper =	BODY BIB;
Level 1 list	REFS =	ref;
Level 2 seq	ref =	auth-primary ?auth-others ref-source PERIOD;
Level 3 seq seq	auth-primary = auth-others =	auth-name-last-first ?PERIOD; COMMA-AND AUTH-SUBS ?PERIOD;
Level 4 list choice	AUTH-SUBS = ref-source =	auth-sub COMMA-AND; TITLE-JOURNAL . . ;
Level 5 struct	TITLE-JOURNAL =	<tml:frag V-YEAR> [.]\s* <tml:frag V-TITLE-PAPER> <tml:frag V-TITLE> <tml:frag END-REF> year letter title journal;

Table A.2: CiteScan warning showing uncited bibliographical reference to “[13]” in Hu & Liu (2004).

** uncited bib reference: `[13] 2003 LEARNING TO CLASSIFY DOCUMENTS ACCORDING TO GENRE`
[^\n][\[]\s*13\s*[\]]
[^\n][\[]\s*13[\s,][0-9,\s]*[\]]
[^\n][\[][0-9,e.g\s]*[,]\s*13\s*[,][\s,][0-9,\s]+[\]]
[^\n][\[][0-9,e.g\s]*[,]\s*13\s*[\]]

Table A.3: The generated citation scan regular expressions for APA style citations to “Webb, E. J., Campbell, D. T., Schwartz, R. D., & Sechrest, L. (1966). ”Unobtrusive measures: Nonreactive research in the social sciences.” Chicago: Rand Mc- Nally.” within Chaffee (1991).

<table>
<tr><td>[\\(\\[]\\s*(?:(?:see\\s*(?:[,]\\s*|\\s+)(?:(?:also|for\\s+example)[,\\s]|e[.]?g[.][,]?)?)|e[.]?g[.]\\s*[,]?|cf[.]\\s*[,]?)?\\s*webb\\s*(?:[,])?\\s*campbell\\s*(?:[,])?\\s*schwartz\\s*(?:[,])?\\s*(?:\\s*[&]\\s*|\\s+and\\s+)sechrest\\s*(?:[,])?\\s*1966(?:[:]\\d+\\s*[fp]+[.]?)?\\s*[\\)\\]]</td></tr>
<tr><td>(?:(?:[\\[\\(;]\\s*(?:(?:see\\s*(?:[,]\\s*|\\s+)(?:(?:also|for\\s+example)[,\\s]|e[.]?g[.][,]?)?)|e[.]?g[.]\\s*[,]?|cf[.]\\s*[,]?)?\\s*webb\\s+et\\s+al\\s*[.]?[,]?\\s*(?:[,])?\\s*1966(?:[:]\\d+\\s*[fp]+[.]?)?\\s*[;])|(?:[;]\\s*webb\\s+et\\s+al\\s*[.]?[,]?\\s*(?:[,])?\\s*1966(?:[:]\\d+\\s*[fp]+[.]?)?\\s*[\\)\\]]))\\bwebb\\s+et\\s+al\\s*[.]?[,]?\\s*(?:[‘][s]?)?\\s*[\\(]\\s*1966(?:[:]\\d+\\s*[fp]+[.]?)?\\s*[\\)]</td></tr>
<tr><td>(?:(?:[\\[\\(;]\\s*(?:(?:see\\s*(?:[,]\\s*|\\s+)(?:(?:also|for\\s+example)[,\\s]|e[.]?g[.][,]?)?)|e[.]?g[.]\\s*[,]?|cf[.]\\s*[,]?)?\\s*webb\\s*(?:[,])?\\s*campbell\\s*(?:[,])?\\s*schwartz\\s*(?:[,])?\\s*(?:\\s*[&]\\s*|\\s+and\\s+)sechrest\\s*(?:[,])?\\s*1966(?:[:]\\d+\\s*[fp]+[.]?)?\\s*[;])|(?:[;]\\s*webb\\s*(?:[,])?\\s*campbell\\s*(?:[,])?\\s*schwartz\\s*(?:[,])?\\s*(?:\\s*[&]\\s*|\\s+and\\s+)sechrest\\s*(?:[,])?\\s*1966(?:[:]\\d+\\s*[fp]+[.]?)?\\s*[\\)\\]]))</td></tr>
<tr><td>\\bwebb\\s*(?:[,])?\\s*campbell\\s*(?:[,])?\\s*schwartz\\s*(?:[,])?\\s*(?:\\s*[&]\\s*|\\s+and\\s+)sechrest\\s*(?:[‘][s]?)?\\s*[\\(]\\s*1966(?:[:]\\d+\\s*[fp]+[.]?)?\\s*[\\)]</td></tr>
<tr><td>[\\(\\[]\\s*(?:(?:see\\s*(?:[,]\\s*|\\s+)(?:(?:also|for\\s+example)[,\\s]|e[.]?g[.][,]?)?)|e[.]?g[.]\\s*[,]?|cf[.]\\s*[,]?)?\\s*webb\\s+et\\s+al\\s*[.]?[,]?\\s*(?:[,])?\\s*1966(?:[:]\\d+\\s*[fp]+[.]?)?\\s*[\\)\\]]</td></tr>
</table>

Table A.4: The 10 most influential papers referenced within Liu (2012).

** PAPER INFLUENCE **
1. 49142 [25] MINING AND SUMMARIZING CUSTOMER REVIEWS (2004) : hu-m, liu-b
2. 19495 [10] MINING COMPARATIVE SENTENCES AND RELATIONS (2006) : jindal-n, liu-b
3. 19326 [10] THUMBS UP OR THUMBS DOWN?: SEMANTIC ORIENTATION APPLIED TO UNSUPERVISED CLASSIFICATION OF REVIEWS (2002) : turney-p
4. 15495 [10] OPINION OBSERVER: ANALYZING AND COMPARING OPINIONS ON THE WEB (2005) : liu-b, hu-m, cheng-j
5. 13923 [7] SENTIMENT ANALYSIS AND SUBJECTIVITY, IN HANDBOOK OF NATURAL LANGUAGE PROCESSING, SECOND EDITION, N (2010) : liu-b
6. 13600 [7] DETERMINING THE SENTIMENT OF OPINIONS (2004) : kim-s, hovy-e
7. 13105 [7] A HOLISTIC LEXICON-BASED APPROACH TO OPINION MINING (2008) : ding-x, liu-b, yu-p
8. 12935 [8] OPINION SPAM AND ANALYSIS (2008) : jindal-n, liu-b
9. 12412 [6] MINING OPINIONS IN COMPARATIVE SENTENCES (2008) : ganapathibhotla-m, liu-b
10. 11466 [5] IDENTIFYING NOUN PRODUCT FEATURES THAT IMPLY OPINIONS (2011) : zhang-l, liu-b

Table A.5: The 10 most influential authors referenced within Hu & Liu (2004).

** AUTHOR INFLUENCE **		
1.	2308 [5]	liu-b: 2004(4), 1998(1)
2.	2177 [6]	wiebe-j: 2003(1), 2000(4), 1999(1)
3.	1948 [5]	fellbaum-c: 1998(2), 1990(3)
4.	1805 [4]	hu-m: 2004(4)
5.	1607 [4]	turney-p: 2002(4)
6.	1232 [3]	hatzivassiloglou-v: 2000(2), 1997(1)
7.	1184 [3]	miller-k: 1990(3)
8.	1184 [3]	gross-d: 1990(3)
9.	1184 [3]	beckwith-r: 1990(3)
10.	1184 [3]	miller-g: 1990(3)

Table A.6: CiteScan query results for ~500 character context around occurrences of the unigram "OPINION" where the unigram "TYPES" is found within 5 words before or after.

> QUERY "opinion" 500 "types" 5
opinion `liu2012sentiment` 001.584 > `orders aspects and their corresponding sentences based on a coherence measure, which tries to optimize the ordering so that they best follow the sequences of aspect appearances in their original postings. Ku, Liang, and Chen (2006) performed blog OPINION summarization, and produced two TYPES of summaries: brief and detailed summaries, based on extracted topics (aspects) and sentiments on the topics. For the brief summary, their method picks up the document/article with the largest number of ...`
opinion `indurkhya2012handbook` 001.718 > `there are also OPINION phrases and idioms, 642 Handbook of Natural Language Processing e.g., cost someone an arm and a leg. Collectively, they are called the OPINION lexicon. They are instrumental for sentiment analysis for obvious reasons. OPINION words can, in fact, be divided into two TYPES, the base type and the comparative type. All the examples above are of the base type. OPINION words of the comparative type are used to express comparative and superlative opinions. Examples of such ...`
...

APPENDIX B. QMS PORTFOLIO DOCUMENT WORD PICTURES

The following figures are word saliency maps for the documents included in the QMS portfolio for Chapter 3.

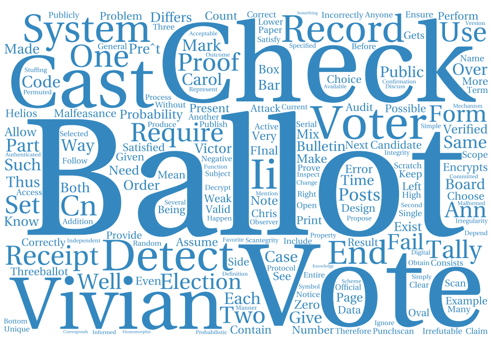


Figure B.1: *Word frequency map for Popoveniuc et al. (2010).*

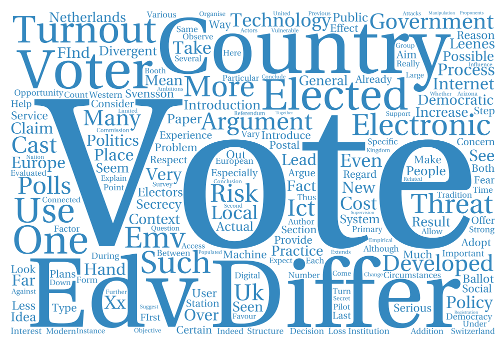


Figure B.2: *Word frequency map for Svensson & Leenes (2003).*

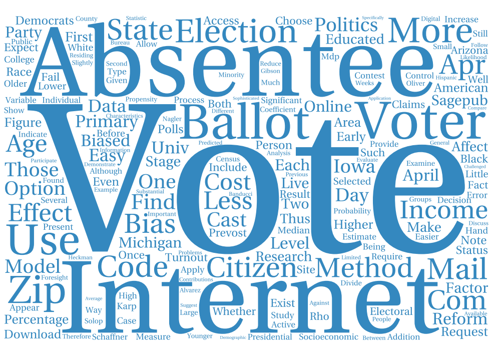


Figure B.3: *Word frequency map for Provost et al. (1998).*

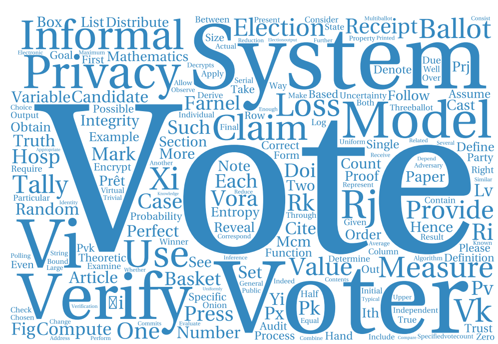


Figure B.4: *Word frequency map for Hosp & Vora (2008).*

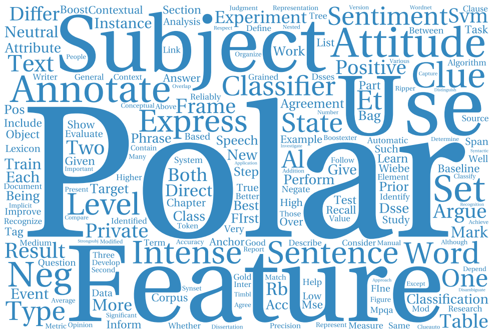


Figure B.5: *Word frequency map for Wilson (2008).*

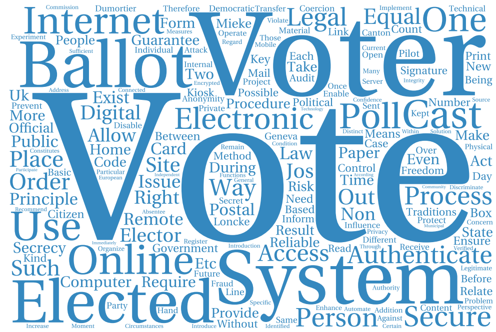


Figure B.6: *Word frequency map for Loncke & Dumortier (2004).*

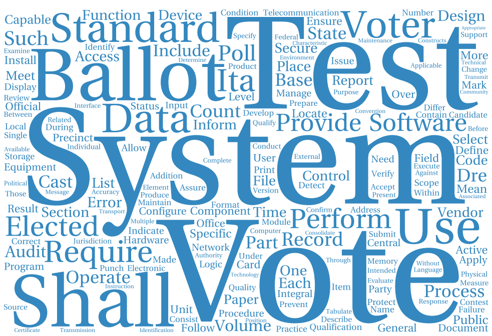


Figure B.7: *Word frequency map for NASED (2002).*

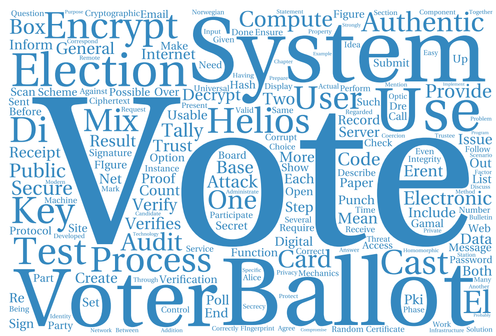


Figure B.8: *Word frequency map for Stenbro (2010).*

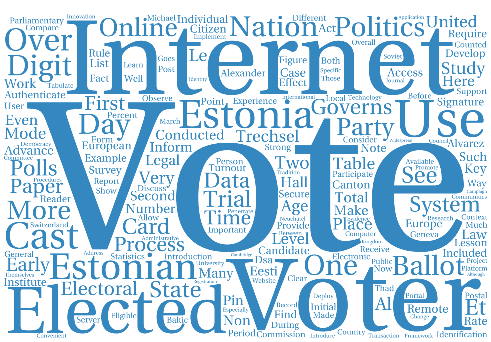


Figure B.9: *Word frequency map for Alvarez et al. (2008).*

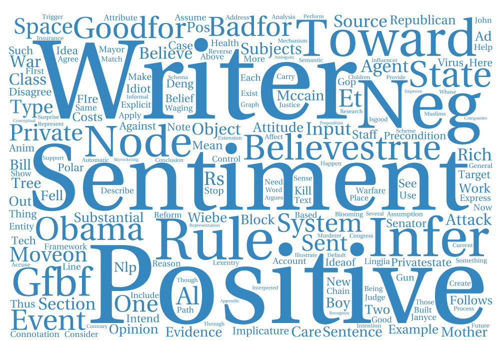


Figure B.10: *Word frequency map for Wiebe & Deng (2014).*

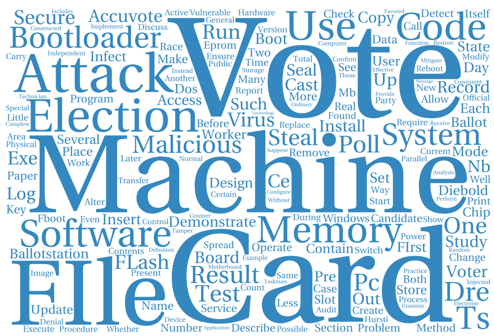


Figure B.11: *Word frequency map for Feldman & Benaloh (2009).*

Figure B.12: *Word frequency map for Somasundaran (2010).*

Figure B.13: *Word frequency map for Ding et al. (2008).*

Figure B.14: *Word frequency map for Bethard et al. (2004).*

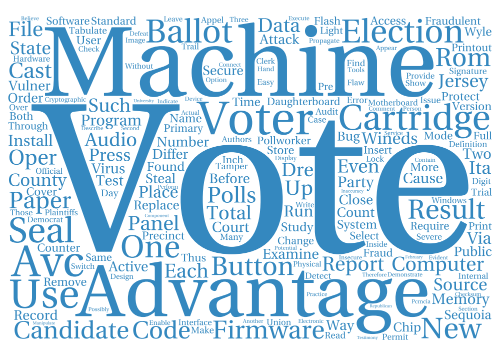


Figure B.15: *Word frequency map for Appel et al. (2009).*

Figure B.16: *Word frequency map for Liu (2012).*

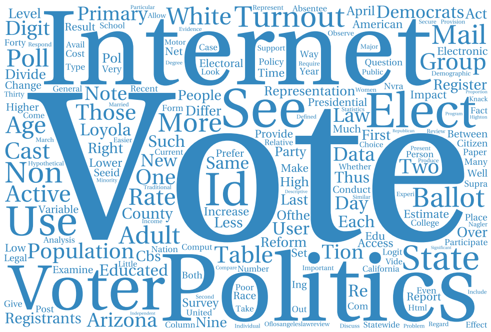


Figure B.17: *Word frequency map for Alvarez & Nagler (2000).*

Figure B.18: *Word frequency map for Zhang & Liu (2011).*

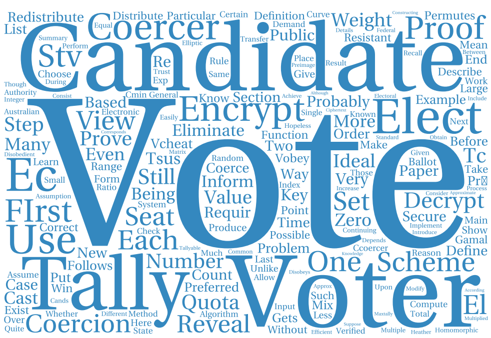


Figure B.19: *Word frequency map for Teague et al. (2008).*

Figure B.20: *Word frequency map for Xu et al. (2007).*

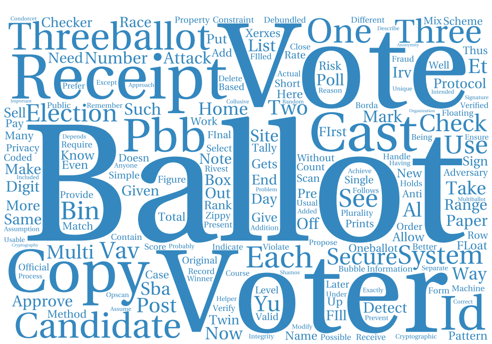


Figure B.21: *Word frequency map for Rivest & Smith (2007).*

Figure B.22: *Word frequency map for Stark (2010).*

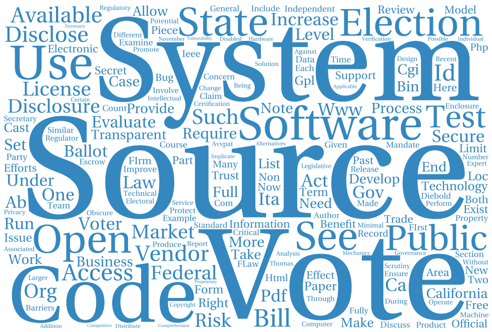


Figure B.23: *Word frequency map for Hall (2006).*

Figure B.24: *Word frequency map for Akkaya (2013).*

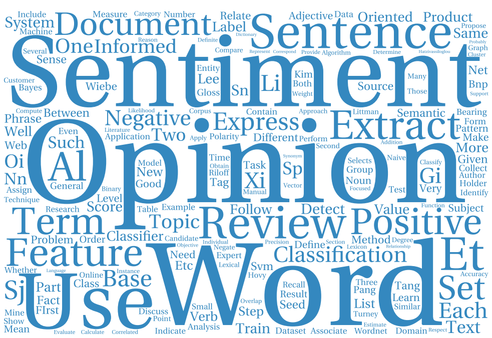


Figure B.25: *Word frequency map for Tang et al. (2009).*

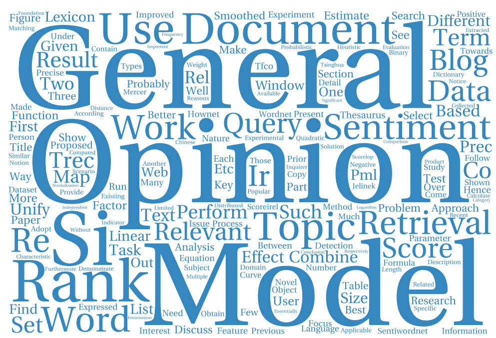


Figure B.26: *Word frequency map for Zhang & Ye (2008).*

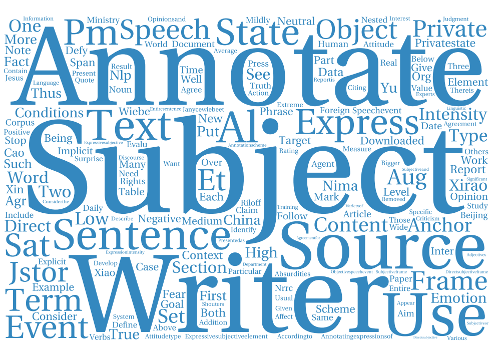


Figure B.27: *Word frequency map for Wiebe et al. (2005).*

Figure B.28: *Word frequency map for Luskin & Fishkin (2005).*

APPENDIX C. THE UNIVERSAL VOTING MARKUP LANGUAGE

The following is the UVML language specification. The full specification can be found at Phillips (2011).

ABNF Rule Definition

```
ballot = [text] 1*(vote [text]) [signature] [text]
vote = rating / ranking
rating = HASH target score
ranking = 2HASH contest selections
signature = 3HASH profile
;
;; rating
target = name / this
score = [undecided] valuation
name = tag *( PERIOD tag )
this = T H I S
;
;; ranking
contest = name
selections = 1*25selection
selection = *HWS [rank] HASH name [[undecided]
valuation]
rank = DIGIT / "1" DIGIT / "2" ZEROTOFIVE
;
;; score
undecided = QUESTION
valuation = quality / importance / outlook
valuation =/ support-opposition / likelihood
;
;; signature
profile = [age [HYPHEN]] gender [jurisdiction]
age = 1*3DIGIT
```

```
gender = iso-5218-gender
jurisdiction = [country] [region] [area-code]
country = [HYPHEN] iso-3166-country
region = [HYPHEN] iso-3166-subdivision
area-code = [HYPHEN] 3*4DIGIT
;
;; name
tag = name-begin *70name-inner name-end
name-begin = LETTER / UPPER-ASCII / UNICODE
name-inner = name-begin / DIGIT
name-end = name-inner
;
;; valuation
quality = among-the-very-best / very-good
quality =/ good / fair / poor
quality =/ very-poor / among-the-very-worst
;
importance = highest-importance / very-important
importance =/ important / unimportant / irrelevant
;
outlook = never-more-optimistic / very-optimistic
outlook =/ optimistic / pessimistic
outlook =/ very-pessimistic / never-more-pessimistic
;
support-opposition = strongly-support / support
support-opposition =/ somewhat-support / somewhat-oppose
support-opposition =/ oppose / strongly-oppose
;
likelihood = definitely / very-likely / likely
likelihood =/ unlikely / very-unlikely / definitely-not
;
;; ~~ quality ~~
among-the-very-best = 5*10STAR
very-good = 4STAR
good = 3STAR
fair = 2STAR
poor = 1STAR
very-poor = STAR MINUS
among-the-very-worst = STAR 2*9MINUS
;
;; ~~ importance ~~
highest-importance = 3*10BANG
very-important = 2BANG
important = BANG
unimportant = BANG MINUS
irrelevant = BANG 2*9MINUS
;
```

```
;; ~~ outlook ~~
never-more-optimistic = 3*10CURRENCY
very-optimistic = 2CURRENCY
optimistic = CURRENCY
pessimistic = CURRENCY MINUS
very-pessimistic = CURRENCY 2MINUS
never-more-pessimistic = CURRENCY 3*9MINUS
CURRENCY = DOLLAR / EURO / POUND / YUAN-YEN
;
;; ~~ support ~~
strongly-support = 3*10PLUS
support = 2PLUS
somewhat-support = PLUS
somewhat-oppose = MINUS
oppose = 2MINUS
strongly-oppose = 3*10MINUS
;
;; ~~ likelihood ~~
definitely = 3*10PERCENT
very-likely = 2PERCENT
likely = PERCENT
unlikely = PERCENT MINUS
very-unlikely = PERCENT 2MINUS
definitely-not = PERCENT 3*9MINUS
;
;; ISO code sets
iso-5218-gender = male / female
iso-3166-country = 2LETTER
iso-3166-subdivision = 1*2DIGIT / 2*3LETTER
male = M
female = F
;
;; symbols
BANG = %x21
HASH = %x23
DOLLAR = %x24
PERCENT = %x25
AMPERSAND = %x26
APOSTROPHE = %x27
STAR = %x2A
PLUS = %x2B
MINUS = %x2D
HYPHEN = %x2D
PERIOD = %x2E
SLASH = %x2F
QUESTION = %x3F
UNDERSCORE = %x5F
```

```
EURO = %x80
POUND = %xA3
YUAN-YEN = %xA5
M = "M" / "m"
F = "F" / "f"
T = "T" / "t"
H = "H" / "h"
I = "I" / "i"
S = "S" / "s"
;
;; symbol groups
LETTERDIGIT = LETTER / DIGIT
LETTER = "A" / "B" / "C" / "D" / "E" / "F" / "G"
LETTER =/ "H" / "I" / "J" / "K" / "L" / "M" / "N"
LETTER =/ "O" / "P" / "Q" / "R" / "S" / "T" / "U"
LETTER =/ "V" / "W" / "X" / "Y" / "Z"
LETTER =/ "a" / "b" / "c" / "d" / "e" / "f" / "g"
LETTER =/ "h" / "i" / "j" / "k" / "l" / "m" / "n"
LETTER =/ "o" / "p" / "q" / "r" / "s" / "t" / "u"
LETTER =/ "v" / "w" / "x" / "y" / "z"
ONETONINE = "1" / "2" / "3" / "4" / "5" / "6" / "7" /"8" / "9"
ZEROTOFIVE = "0" / "1" / "2" / "3" / "4" / "5"
UPPER-ASCII = %xC0-FF
UNICODE = PLANE0
PLANE0 = %x0100-D7FF / %xE000-FDCF
PLANE0 =/ %xFDF0-FFFD
;; NOTE: java/scala
lack support for
PLANE1-2
;; PLANE1 = %x10000-1FFFD
;; PLANE2 = %x20000-2FFFD
```

BIBLIOGRAPHY

Akkaya, C. (2013). *SUBJECTIVITY WORD SENSE DISAMBIGUATION: A TOOL FOR SENSE-AWARE SUBJECTIVITY ANALYSIS*. PhD thesis, University of Pittsburgh.

Albig, W. (1957). Two decades of opinion study: 1936-1956. *Public Opinion Quarterly*, *21*(1), 14.

Alvarez, M., Hall, T., & Treschsel, A. (2008). Internet voting in estonia. Technical report, VTP Working Paper.

Alvarez, R. M. & Nagler, J. (2000). Likely consequences of internet voting for political representation, the. *Loy. LAL Rev.*, *34*, 1115.

Appel, A., Ginsburg, M., Hursti, H., Kernighan, B., Richards, C., Tan, G., & Venetis, P. (2009). The New Jersey voting-machine lawsuit and the AVC Advantage DRE voting machine. In *Proceedings of the 2009 conference on Electronic voting technology / workshop on trustworthy elections*, (pp. 5–5). USENIX Association.

Bethard, S., Yu, H., Thornton, A., Hatzivassiloglou, V., & Jurafsky, D. (2004). Automatic extraction of opinion propositions and their holders. In *2004 AAAI Spring Symposium on Exploring Attitude and Affect in Text*, (pp. 2224).

Brill, E. (1993). *A corpus-based approach to language learning*. PhD thesis, Citeseer.

Chaffee, S. H. (1991). *Explication*, volume 1. Sage Publications, Incorporated.

De Montaigne, M. (1580). *The complete essays of Montaigne*, volume 1.

Ding, X., Liu, B., & Yu, P. S. (2008). A holistic lexicon-based approach to opinion mining. In *Proceedings of the 2008 International Conference on Web Search and Data Mining*, (pp. 231–240). ACM.

Feldman, A. & Benaloh, J. (2009). On subliminal channels in encrypt-on-cast voting systems. In *Proceedings of the 2009 conference on Electronic voting technology/workshop on trustworthy elections*, (pp. 12–12). USENIX Association.

Fishkin, J. (2000). Virtual democratic possibilities: Prospects for Internet democracy. In *conference Internet, Democracy and Public Goods. Belo Horizonte, Brazil*.

Fishkin, J., He, B., Luskin, R., & Siu, A. (2006). Deliberative democracy in an unlikely place: deliberative polling in China. *British journal of political science*, *1*(-1), 1–14.

Fishkin, J., He, B., & Siu, A. (2008). Public consultation through deliberation in China: the first Chinese deliberative poll. *Governance reform under real-world conditions: citizens, stakeholders, and voice*, 461.

Fishkin, J. & Luskin, R. (2005). Experimenting with a democratic ideal: Deliberative polling and public opinion. *Acta Politica*, *40*(3), 284–298.

Fishkin, J., Luskin, R., & Jowell, R. (2000). Deliberative polling and public consultation. *Parliamentary Affairs*, *53*(4), 657–666.

Gallup, G. (1957). The changing climate for public opinion research. *Public Opinion Quarterly*, *21*(1), 23.

Hall, J. (2006). Transparency and access to source code in electronic voting. In *Proceedings of the USENIX/Accurate Electronic Voting Technology Workshop 2006 on Electronic Voting Technology Workshop*, (pp. 8–8). USENIX Association.

Hosp, B. & Vora, P. (2008). An information-theoretic model of voting systems. *Mathematical and Computer Modelling*, *48*(9-10), 1628–1645.

Hu, M. & Liu, B. (2004). Mining and summarizing customer reviews. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, (pp. 168–177). ACM.

Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis Lectures on Human Language Technologies*, *5*(1), 1–167.

Loncke, M. & Dumortier, J. (2004). Online voting: a legal perspective. *International Review of Law, Computers &# 38; Technology*, *18*(1), 59–79.

Luskin, R. & Fishkin, J. (2005). Deliberative polling, public opinion, and democracy: the case of the National Issues Convention. In *Slightly revised from a paper presented at the annual meeting of the American Political Science Association, Boston, MA, September 2-6, 1998*.

Luskin, R., Fishkin, J., & Plane, D. (1999). Deliberative polling and policy outcomes: electric utility issues in Texas. In *annual meeting of the Midwest Political Science Association, Chicago*, volume 4, (pp. 1999).

NASED, N. S. E. D. A. (2002). Voting system standards, volume 1.

Phillips, E. (2011). The universal voting markup language (uvml). In *IETF, DRAFT RFC*.

Popoveniuc, S., Kelsey, J., Regenscheid, A., & Vora, P. (2010). Performance requirements for end-to-end verifiable elections. In *Proceedings of the 2010 international conference on Electronic voting technology/workshop on trustworthy elections*, (pp. 1–16). USENIX Association.

Provost, F. J., Fawcett, T., & Kohavi, R. (1998). The case against accuracy estimation for comparing induction algorithms. In *ICML*, volume 98, (pp. 445–453).

Rivest, R. & Smith, W. (2007). Three voting protocols: ThreeBallot, VAV, and Twin. In *Proceedings of the USENIX Workshop on Accurate Electronic Voting Technology*, (pp. 16–16). USENIX Association.

Somasundaran, S. (2010). *Discourse-level relations for Opinion Analysis*. PhD thesis, University of Pittsburgh.

Stark, P. B. (2010). Super-simple simultaneous single-ballot risk-limiting audits. In *Proceedings of the 2010 Electronic Voting Technology Workshop/Workshop on Trustworthy Elections (EVT/WOTE'10). USENIX*.

Stenbro, M. (2010). A survey of modern electronic voting technologies.

Svensson, J. & Leenes, R. (2003). E-voting in Europe: Divergent democratic practice. *Information Polity*, *8*(1, 2), 3–15.

Tang, H., Tan, S., & Cheng, X. (2009). A survey on sentiment detection of reviews. *Expert Systems with Applications*, *36*(7), 10760–10773.

Teague, V., Ramchen, K., & Naish, L. (2008). Coercion-resistant tallying for STV voting. In *Proceedings of the conference on Electronic voting technology*, (pp. 1–14). USENIX Association.

Wiebe, J. & Deng, L. (2014). An account of opinion implicatures. *CoRR, abs/1404.6491*.

Wiebe, J., Wilson, T., & Cardie, C. (2005). Annotating expressions of opinions and emotions in language. *Language resources and evaluation*, *39*(2-3), 165–210.

Wilson, T. A. (2008). *Fine-grained subjectivity and sentiment analysis: recognizing the intensity, polarity, and attitudes of private states*. ProQuest.

Xu, R., Wong, K.-F., & Xia, Y. (2007). Opinmine–opinion analysis system by cuhk for ntcir-6 pilot task. In *Proceedings of the 6th NTCIR Workshop*.

Zhang, L. & Liu, B. (2011). Identifying noun product features that imply opinions. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: short papers-Volume 2*, (pp. 575–580). Association for Computational Linguistics.

Zhang, M. & Ye, X. (2008). A generation model to unify topic relevance and lexicon-based sentiment for opinion retrieval. In *Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval*, (pp. 411–418). ACM.